BROOKINGS
www.brookings.edu

ISBN 978-0-8157-2870-2
90000>
9 780815 728702

Brookings Papers
ON ECONOMIC ACTIVITY

BPEA
FALL
2015

BROOKINGS

David H. Romer and Justin Wolfers, Editors

Brookings Papers
ON ECONOMIC ACTIVITY

FALL 2015

Brookings Papers
ON ECONOMIC ACTIVITY

FALL 2015

DAVID H. ROMER
JUSTIN WOLFERS
Editors

BROOKINGS INSTITUTION PRESS
Washington, D.C.

ISSN 0007-2303
ISBN: Print 978-0-8157-2870-2; epub: 978-0-8157-2871-9

Brookings Papers
ON ECONOMIC ACTIVITY

FALL 2015

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

The conference and the journal are based upon the work partially supported by the National Science Foundation under Grant No. 1459089.

The papers and discussant remarks reflect the views of the authors and not necessarily the views of the funding organizations or the staff members, officers, or trustees of the Brookings Institution.

CALL FOR PAPERS Most papers that appear in the *Brookings Papers on Economic Activity* are solicited by the editors, but the editors also consider submitted proposals. Editorial decisions are generally made about a year in advance of each conference. Therefore, proposals should be received by September 1 for the following fall conference and by April 1 for the following spring conference. Proposals should be no more than three double-spaced pages and should be sent to bpea@brookings.edu or to Sarah Holmes, Brookings Institution, 1775 Massachusetts Ave., NW, Washington, D.C. 20036.

ACCESSING THE JOURNAL For information about previous issues of the journal, participants in this conference, and agendas for upcoming conferences, visit www.brookings.edu/about/projects/bpea. 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 OCLC subscription department at 800-848-5878) and Project Muse (http://muse.jhu.edu). Archived issues of the *Brookings Papers on Economic Activity* are available through JSTOR (www.jstor.org).

EDITORS, AUTHORS, DISCUSSANTS, PANEL ADVISERS, AND STAFF FOR THE ONE-HUNDREDTH CONFERENCE

Henry Aaron *Brookings Institution*
Andrew Abel *University of Pennsylvania*
Katharine Abraham *University of Maryland*
Hassan Afrouzi *University of Texas, Austin*
Elizabeth Akers *Brookings Institution*
Martin Baily *Brookings Institution*
Sandra Black *Council of Economic Advisers*
Alan S. Blinder *Princeton University*
Michael Boldin *Federal Reserve Bank of Philadelphia*
William Brainard *Yale University*
Christopher Carroll *Consumer Financial Protection Bureau*
Gerald Cohen *U.S. Department of the Treasury*
Olivier Coibion *University of Texas, Austin*
Richard Cooper *Harvard University*
Steven Davis *University of Chicago*
J. Bradford DeLong *University of California, Berkeley*
Jane Dokko *Brookings Institution*
Jennifer Doleac *Brookings Institution*
John C. Driscoll *Board of Governors of the Federal Reserve System*
Karen Dynan *U.S. Department of the Treasury*
Janice Eberly *Northwestern University*
Wendy Edelberg *Congressional Budget Office*
Douglas J. Elliott *Brookings Institution*
Douglas W. Elmendorf *Brookings Institution*
William B. English *Board of Governors of the Federal Reserve System*
Bruce Fallick *Federal Reserve Bank of Cleveland*
Martin Feldstein *Harvard University*
Jason Furman *Council of Economic Advisers*
Joseph E. Gagnon *Peterson Institute for International Economics*
William G. Gale *Brookings Institution*
Ted Gayer *Brookings Institution*
Sarena F. Goodman *Board of Governors of the Federal Reserve System*
Robert Gordon *Northwestern University*
Tracy Gordon *Urban Institute*
Yuriy Gorodnichenko *University of California, Berkeley*
Joshua Gotbaum *Brookings Institution*
Carol Graham *Brookings Institution*
Mark Griffiths *International Monetary Fund*
Erica Groshen *U.S. Bureau of Labor Statistics*
Simona M. Hannon *Board of Governors of the Federal Reserve System*
Joshua K. Hausman *University of Michigan*
Christopher L. House *University of Michigan*
Caroline Hoxby *Stanford University*
Glenn Hubbard *Columbia University*
Yannis Ioannides *Tufts University*
Michael Kiley *Board of Governors of the Federal Reserve System*
Jeffrey R. Kling *Congressional Budget Office*
Donald Kohn *Brookings Institution*
Saten Kumar *Auckland University of Technology*
John Lipsky *Johns Hopkins University*
Adam Looney *U.S. Department of the Treasury*
Annamaria Lusardi *George Washington University*
N. Gregory Mankiw *Harvard University*
Donald Marron *Urban Institute*

Editors' Introduction

THE BROOKINGS PANEL ON ECONOMIC ACTIVITY held its hundredth conference in Washington, D.C., on September 10 and 11, 2015, and this volume is the hundredth issue of the *Brookings Papers on Economic Activity.* From the very first meeting, held in April 1970, the Brookings Panel and *Brookings Papers* have been committed to rigorous, empirically driven analyses of issues directly relevant to important questions for macroeconomic policy. But the differences between this volume and the first one show how much both *Brookings Papers* and the field of macroeconomics have changed over the past 45 years.

All the contributions in the first volume were concerned with short-run macroeconomic fluctuations in the United States. They all relied on relatively aggregate data, for example at the level of major industries or demographic groups, at frequencies ranging from roughly monthly to annual. No contributions used data from outside the United States or from before 1950, and the empirical methods consisted almost entirely of summary statistics and linear regressions. The 16 contributors came from 10 institutions, all of them in the United States. Only one contributor (and only one of the 30 meeting participants) was female.

In the current volume, there are contributions devoted to understanding recent developments in Greece, Japan, and Portugal, and another that uses evidence drawn mainly from New Zealand. Two other contributions focus on the growth of student loan debt and rules for assessing the budgetary impact of legislation. Only one is devoted to an issue closely related to short-run macroeconomic fluctuations in the United States. The data used in this volume include responses to original surveys, a newly created individual-level data set from merged administrative sources, prices of sophisticated financial instruments, daily weather information from a large number of U.S. metropolitan areas, and recently collected information on

Greek bond issuance extending back to the early nineteenth century. And while simple facts and summary statistics continue to play a central role in much of the analysis, other empirical approaches include nonlinear time-series techniques, calibration and simulation of dynamic general equilibrium models, discrete choice models, and qualitative analyses of particular institutions and episodes. The 30 contributors are drawn from 24 institutions, and seven contributors are based outside the United States. Seven of the 30 contributors are female.

IN THE FIRST PAPER OF THIS VOLUME, Adam Looney and Constantine Yannelis study the explosive growth of federal student loan debt and student loan defaults using an important new data set constructed from administrative sources, with 46 million annual observations on 4 million borrowers. The data set includes information about loans, repayments, and defaults and other types of nonpayment, as well as earnings and income. A central finding is that to the extent there is any crisis in student loans, it is concentrated among individuals who borrowed to attend two types of schools that traditionally played only a small role in federal student loans: for-profit schools and, to a lesser extent, 2-year schools. Despite the Great Recession and its aftermath, default rates by traditional borrowers have remained moderate; the large rise in overall defaults stems from the increased share of nontraditional borrowers and their soaring default rates. Looney and Yannelis show that many of the schools that are the source of the largest amounts of borrowing in recent years are for-profits whose borrowers have extremely high rates of default and who on the whole are making essentially no progress (or even negative progress) in paying off their loans. But they project that recent shifts away from nontraditional borrowers, improved labor-market outcomes, and changes in the student loan program are likely to reduce default rates going forward. As Looney and Yannelis stress, their findings paint a quite different and more nuanced picture than the simple view of a widespread student debt crisis across all types of borrowers, and their findings raise important questions for policymakers.

Traditional estimates of the budgetary effects of proposed legislation ignore any effects operating through the legislation's likely impact on overall economic activity; in the jargon of budget analysts, the traditional approach involves "static" rather than "dynamic" scoring. In the second paper, Douglas Elmendorf lays out the case that not doing dynamic scoring is no more appropriate than ignoring microeconomic reallocations that proposed legislation would likely cause, which have

long been considered in budgetary assessments. He also argues that the agencies that estimate the budgetary impacts of proposed legislation in Congress do have the resources and technical skills needed to carry out reasonable dynamic scoring, at least for most major legislation, and that the usual objections to dynamic scoring are ill-founded or no longer relevant. Finally, he argues that the requirements for limited dynamic scoring recently adopted by Congress should be modified to have greater symmetry in the treatment of tax and spending changes and in the handling of the two major parties.

Inflation targeting, which in some form is now practiced by most central banks, is often thought to anchor price-setters' inflation expectations. In the third paper of this volume, Saten Kumar, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko challenge this view using the results of a unique survey of managers in New Zealand, the country with the longest and best-established history of inflation targeting. The managers report that information about inflation is potentially important to their business decisions. But the survey also shows that managers know little about the central bank or its objectives; that they pay little attention to the central bank; that their beliefs about recent inflation are often highly inaccurate; and, crucially, that individual managers' point estimates of future inflation often depart widely from the central bank's target and that they have high levels of uncertainty about inflation. These results are difficult to square with either standard views of inflation targeting or standard models of rational agents, so they represent an important challenge for both policymakers and economists.

The fourth paper, by Michael Boldin and Jonathan Wright, fits firmly into *Brookings Papers*' long-standing concern with the nitty-gritty of macroeconomic data. Boldin and Wright focus in on the fact that although statistical agencies routinely adjust data for the effects of usual seasonal weather patterns and other regular sources of seasonal variation, there is no comparable treatment of unusual weather. Yet for many purposes, policymakers and analysts want to know how reported data have been affected by unusual weather. Boldin and Wright therefore propose and implement a method for adjusting the most widely watched monthly macroeconomic indicator—payroll employment—for unusual weather effects. They find that those effects are often substantial and can have important implications for the interpretation of the data.

One of the most significant and frightening developments in the world economy over the past few years has been the economic, political, and social turmoil in Greece. A portion of the Fall meeting was therefore

devoted to a special symposium where four author teams presented analyses of different aspects of the Greek crisis. Julian Schumacher and Beatrice Weder di Mauro focus on Greece's debt. They show that, because most of the lending to Greece is from official lenders at below-market rates, the present value of the country's debts is far less than its face value. They also address the costs and benefits of the relaxation of traditional criteria for official lending that took place as policymakers struggled to deal with the crisis. Carmen Reinhart and Christoph Trebesch place the current crisis in the broader context of Greek history, showing that it is just one of multiple episodes where heavy external borrowing has led to disaster. Christopher House and Linda Tesar, in the spirit of the dynamic scoring that is the subject of Elmendorf's paper, focus on macroeconomic feedbacks from adjustment policies. Their analysis implies that those feedbacks are likely to cause the actual budgetary benefits of tax increases or spending reductions to be far less than is implied by calculations that ignore the feedbacks. Finally, Yannis Ioannides and Christopher Pissarides argue that the emphasis on austerity and comprehensive structural reform is misplaced, and that the central near-term focus of efforts to heal Greece's economy should be on reforms of the product market.

The final two papers are updates of analyses that appeared in earlier volumes. In the first, Joshua Hausman and Johannes Wieland revisit their work from the Spring 2014 volume concerning "Abenomics," as the recent set of economic policies in Japan is known, with emphasis on its monetary policy component. They find that the evidence continues to support their earlier conclusion that while Abenomics has not led to a dramatic change in the course of the Japanese economy, it has had moderate benefits and few discernible costs. They also highlight two puzzles: the apparently small effects on net exports of the large depreciation of the yen, and the seemingly small response of consumption to the significant fall in the real interest rate. In the second update, Ricardo Reis revisits his analysis of Portugal's economy from the Spring 2013 volume. His thesis is that the conventional wisdom about the Portuguese economy is wrong in two critical ways. First, he finds that despite policymakers' enormous focus on austerity, Portugal has made little or no progress in addressing its long-run fiscal problems. Second, he argues, again contrary to standard views, that Portugal has made significant progress in undertaking structural reforms of its economy.

IN ADDITION TO BEING THE HUNDREDTH VOLUME of the *Brookings Papers on Economic Activity,* this is our last volume as its editors. To

paraphrase an advertising slogan, *Brookings Papers* is never yours; you only take care of it for the next generation. *Brookings Papers* is a treasure that plays a unique and vital role both in economic research and in the interface between economic research and policy. We are honored to have been entrusted with taking care of it for the past seven years, and we are gratified and excited that the editorship will now be in the very talented hands of Janice Eberly and James Stock. We look forward to watching as *Brookings Papers* enters its second "century."

ADAM LOONEY
U.S. Department of the Treasury

CONSTANTINE YANNELIS
New York University, Stern School of Business

A Crisis in Student Loans? How Changes in the Characteristics of Borrowers and in the Institutions They Attended Contributed to Rising Loan Defaults

ABSTRACT This paper examines the rise in student loan default and delinquency. It draws on a unique set of administrative data on federal student borrowing matched to earnings records from de-identified tax records. Most of the increase in default is associated with borrowers at for-profit schools, 2-year institutions, and certain other nonselective institutions. Historically, students at these institutions have constituted a small share of all student borrowers. These nontraditional borrowers have largely come from lower-income families, attended institutions with relatively weak educational outcomes, faced poor labor market outcomes after leaving school, and defaulted at high rates. In contrast, default rates have remained low among borrowers who attended most 4-year public and nonprofit private institutions and among graduate school borrowers—who collectively represent the vast majority of the federal loan portfolio—despite the severe recession and these borrowers' relatively high loan balances. The higher earnings, low rates of unemployment, and greater family resources of this latter category of borrowers appear to have helped them avoid adverse loan outcomes even during times of hardship. Decomposition analysis indicates that changes in the characteristics of borrowers and the institutions they attended are associated with much of the doubling in default rates between 2000 and 2011, with changes in the type of schools attended, debt burdens, and labor market outcomes explaining the largest share.

Between 2000 and 2014, the total volume of outstanding federal student debt nearly quadrupled to surpass $1.1 trillion, the number of student loan borrowers more than doubled to reach 42 million, and default

rates among recent student loan borrowers rose to their highest levels in 20 years. This increase in debt and default and more widespread concern about the effects of student loan debt on young Americans' lives has contributed to a belief that there is a crisis in student loans. Using new administrative data sources, we examine recent changes in the market for federal student loans with a particular focus on the sources of rising default rates, the roles played by educational institutions, and the labor market outcomes of borrowers.

These data show that to the extent that there is a crisis, it is concentrated among borrowers who attended for-profit schools and, to a lesser extent, 2-year institutions and certain other nonselective institutions. We refer to these borrowers as "nontraditional" because, as students, they tend to be older, often enroll less than full time, and are living independently of their parents, and also because historically there were relatively few for-profit students and because 2-year students rarely borrowed. As a result, in 2000 these borrowers represented a small share of all federal student loan borrowers and an even smaller share of loan balances.

However, during and soon after the recession, the number of nontraditional borrowers grew to represent almost half of all new borrowers. They experienced poor labor market outcomes, had few family resources, and owed high debt burdens relative to their earnings. Their default rates skyrocketed. Of all the students who left school, started to repay federal loans in 2011, and had fallen into default by 2013, about 70 percent were nontraditional borrowers.[1]

In contrast, the majority of undergraduate and graduate borrowers from 4-year public and private (nonprofit) institutions, or "traditional borrowers," have experienced strong labor market outcomes and low rates of default, despite having the largest loan balances and facing the severe headwinds of the recent recession. While the number of traditional borrowers also increased rapidly over time, recent borrowers' family backgrounds and labor market outcomes are not much different from their peers' in earlier years, especially for graduate students and undergraduates at relatively selective institutions. In fact, traditional borrowers earned more, on average, in 2013 than their peers had in 2002. While graduates in the late 2000s were hit harder than other cohorts by the recession, the unemployment rate of traditional borrowers who left school and started repaying their loans in

1. Borrowers from nonselective 4-year schools accounted for an additional 12 percent of defaults.

2011 was 7.7 percent in 2013 compared to 6.6 percent for the comparable cohort of recent borrowers in 2002.

These results derive from a new database formed by the merger of administrative records on student loan burdens to earnings information from de-identified tax records. The data provide annual information on student characteristics, the institutions they attended, loan balances and loan status from 1970 to 2014, and labor market outcomes from 1999 to 2013 for a 4-percent sample of all federal student borrowers. The sample includes about 46 million annual observations on 4 million individual borrowers, assembled from hundreds of millions of individual records of loan transactions, aid applications, and earnings records. These data were assembled to improve budget estimates and inform policy regarding programs with both spending and tax components. They also provide unique advantages over prior survey and credit-panel data sets because they allow detailed examination of the role played in student loan defaults by institutions, labor market outcomes, and other potential contributing factors.[2]

These data show that the number of new nontraditional borrowers increased steadily since the mid-1990s, as enrollment in for-profit institutions returned to growth after having declined earlier in the 1990s and then surged during the recession and as the weak labor market boosted enrollment and increased borrowing rates, particularly among 2-year students.[3] Because of the relatively short enrollment durations of many of these new borrowers, the combination of new enrollment and rapid turnover resulted in a flood of nontraditional borrowers, disproportionate to their share of enrollment, who were out of school and into loan repayment after the recession. For instance, in 2011, while for-profit students made up only 9 percent of all postsecondary students (according to the National Center for Education Statistics) and 25 percent of all active federal borrowers, they represented more than 31 percent of borrowers leaving school and starting to repay federal loans that year. Combined with students from 2-year institutions, who represented an additional 16 percent of borrowers starting to repay loans that year, this meant that almost half of borrowers in their first years of repayment were nontraditional borrowers.

2. Detailed tabulations of this database are described in the online appendix. The online appendixes for this and all other papers in this volume may be found at the *Brookings Papers* web page, www.brookings.edu/bpea, under "Past Editions."

3. High default rates in the 1980s led to major accountability changes in the student loan program, which contributed to pushing more than 1,500 for-profit schools out of business and to declines in enrollment and borrowing at for-profit institutions through the early 1990s. This period coincided with a sharp decline in the overall default rate on student loans.

In addition to being more numerous and in their earliest years of loan repayment after the recession, recent nontraditional borrowers appear to be a particularly high-risk population. They tend to be older when they first enroll, to be from lower-income families, and to live in poorer neighborhoods. They are more likely to be first-generation borrowers. They attend programs they are less likely to complete and, after enrollment, are more likely to live in or near poverty and to experience weak labor market outcomes, outcomes that worsened disproportionately during the recession. And their loan burdens, though smaller on average both in absolute terms and relative to their earnings, have tended to increase faster over time.

All these factors contributed to high default rates among nontraditional borrowers. About 30 percent of nontraditional borrowers required to start repayment on loans in 2011 defaulted within three years, compared to 13 percent among traditional undergraduate borrowers and 3 percent among graduate borrowers. Many more appear to be struggling with their loans but have avoided default through protections such as forbearance, deferment, and income-based repayment programs, which allow borrowers to suspend or make reduced payments during times of hardship. Using decomposition analysis, we find that changes in observable characteristics—like the backgrounds of students, their labor market outcomes, and the schools they attend—can explain between half and two-thirds of this increase in default, with changes in the types of institutions attended alone explaining between one-quarter and one-half of the increase in default rates between 2000 and 2011. However, much of the increase in default rates, particularly among nontraditional borrowers, cannot be explained simply by factors like their family background or labor market outcomes, suggesting that factors we cannot observe, such as the quality of the education received, students' satisfaction with their institutions, and other financial or economic difficulties specific to nontraditional borrowers, may also be driving up default rates.

These high rates of default are unlikely to persist because of the recent normalization in enrollment patterns post-recession, increased scrutiny and policing of for-profit institutions, and other factors that have contributed to a decline in the number of nontraditional borrowers. From 2010 to 2014, the number of new borrowers fell by 44 percent at for-profit schools and by 19 percent at 2-year institutions. Because of the relatively long life cycle of a student loan, these changes will not be fully felt for several years.[4]

4. A loan originated to a first-time borrower will not be "eligible" for default until the student completes her educational career, enters a six-month grace period, and then spends about a year in repayment on the loan.

In addition, rising enrollment in income-based repayment programs will help many borrowers experiencing economic hardships avoid default. The decline in the rate of new borrowers and an uptick in the number of borrowers paying off loans have already contributed to a sharp slowdown in the growth rate of borrowers and aggregate debt. In 2014, for instance, the number of borrowers increased by about 1 million, down from an average annual increase of 2 million from 2009 to 2012.

One reason traditional student loan borrowers have avoided default is that they experience favorable labor market outcomes, with low rates of nonemployment (even in the recession) and relatively high earnings; moreover, they are more likely to come from higher-income families in the first place. Most traditional borrowers have not accumulated large balances. While average debt burdens have increased and some borrowers have accumulated very large balances (4 percent of borrowers had balances over $100,000 and 14 percent had balances over $50,000 in 2014), most borrowers with large balances are graduate students, parents, and "independent" undergraduate borrowers often from for-profit schools. Indeed, one consequence of these patterns is that borrowers in the top 20 percent of the income distribution owe more than one-third of outstanding student loan debt.

Beyond examining the sources of the rise in default and delinquency, these data also inform a broader debate regarding the implications of rising student indebtedness. One concern is that rising rates of default reflect excessive borrowing and overextended finances, which could impair students' abilities to finance first homes and to live independently of their families, or could constrain their occupational choices, reducing rates of homeownership and marriage, or their entrepreneurial risk taking.[5] Our results suggest a potentially different interpretation for many of the observed relationships between rising student borrowing and worsening outcomes: a shift in the composition of borrowers toward higher-risk or more disadvantaged individuals. Just as these shifts contribute to higher default rates, they may also contribute to lower rates of homeownership or to constrained occupational choices. In fact, increases in default rates due to compositional shifts could overshadow relatively beneficial investments in higher education, which may be less worrisome or even desirable (Akers and Chingos 2014; Avery and Turner 2012; Dynarski and Kreisman 2013; Sun and Yannelis 2016). Indeed, for most borrowers

5. See, for instance, Baum (2015), Bleemer and others (2014), Field (2009), Gicheva (2013), Gicheva and Thompson (2014), Ionescu (2009, 2011), Marx and Turner (2015), Mezza, Sommer, and Sherlund (2014), and Shao (2014).

(and the majority of the student loan portfolio) the educational investments financed with their loans are associated with favorable economic outcomes, and most borrowers appear able to manage their debt even during recessionary periods.

Ultimately, an important question for understanding the welfare consequences of student aid programs and for developing new policies to improve their effectiveness is whether attending college was worthwhile for borrowers, even those with high rates of default. On average, education is among the most productive of investments individuals can make because the benefits of higher earnings and better well-being accumulate over a lifetime (Greenstone and Looney 2011). But the relatively weak labor market performance, high default rates, and increasing debt burdens of many borrowers raise concerns that not all students are better off. One specific area of concern is that the costs of education and the debts of borrowers have increased relative to labor market returns. Understanding why costs and debt burdens are rising, and parsing out the relative contributions of the recession's effects on enrollment, households' savings and ability to borrow, contraction in public support for education, rising costs of attendance, and other factors, would help identify whether steep increases are associated with changes in the return on students' educational investments.

Similarly, on the other side of the cost-benefit ledger, the benefits of loan-financed education depend on the quality of education provided and the labor market return specific to those investments. That requires knowing not just how borrowers are doing today, but the difficult-to-measure counterfactual of how they would have fared if they had not attended a particular institution. Understanding the differences in the costs and returns to different institutions—and how to encourage higher-return investments—is therefore a key predicate for improving federal loan programs. This appears to be especially true for nontraditional borrowers, for whom educational opportunities appear to vary more in cost and quality, and for whom their educational and financing choices appear to have much larger implications for their longer-term well-being.

The remainder of this paper is organized as follows. In section I we provide background on the structure of federal student loan programs. In section II we discuss the data sources used in the paper. In section III we analyze the factors associated with the increase in student loan debt and discuss the rise of nontraditional borrowers and the implications for borrowing and for borrowers' default and labor market outcomes. In section IV we provide an analysis of the characteristics of nontraditional borrowers and their backgrounds. In section V we provide information on borrowers'

labor market outcomes. In section VI we analyze the debt burdens of borrowers over time. In section VII we focus on a key outcome—loan repayment—and conduct regression and decomposition analysis of factors associated with the rise in student loan default; we then discuss the flows of borrowers during and after the Great Recession as well as the potential implications for future repayment. In section VIII we conclude and provide suggestions for further research.

I. Background: The Structure of Federal Student Loan Programs

The analysis in this paper focuses on federal student lending programs, which were first established in 1958 to provide low-cost loans to students and were subsequently expanded several times, notably under the Higher Education Act of 1965. These federal student lending programs accounted for the nation's largest source of nonmortgage household debt in 2014.[6] The aim of these student loan programs was to alleviate credit constraints for borrowers, who internalize many of the benefits of education.[7]

In 1966, President Lyndon Johnson articulated the purpose of the student loan program in these words:

> Under this new loan program, families will finance college education for their children in the same way that they finance the purchase of a home: through long-term, federally guaranteed private loans. For millions of families, the financial burden of college education will now be lifted; new opportunities will open for American students. (Johnson 1966)

The vast majority of student loans in the United States are federally guaranteed or direct loans made by the Department of Education.[8] The main federal lending program today is the Federal Direct Loan program, which was created by the Higher Education Amendments of 1992. Since

6. The Department of Education provides more information on federal student lending programs at http://federalstudentaid.ed.gov/site/front2back/overview/overview/fb_02_01_0040.htm. It also provides extensive information on student loan interest rates at https://studentaid.ed.gov/sa/types/loans/interest-rates.

7. This stands in contrast to many other countries, where governments finance education expenses. The welfare implications of direct government financing of college costs are controversial, since while there are externalities associated with education, borrowers internalize many of the benefits of higher education. For discussions of financing higher education, see Hartman (1972), Johnson (2006), and Psacharopoulos and Papakonstantinou (2005).

8. Private student loans, which are not included in these data, are a small portion of the aggregate total student loan volume, amounting to less than one-tenth of all student loans disbursed between 2009 and 2013 (College Board 2014).

2010, the Direct Loan program has accounted for all federal student loans. Under this program, postsecondary institutions originate loans under federal lending rules, and loan servicing is handled by the Department of Education through private servicing contractors.

Direct Loans, which can be made both to undergraduate and graduate students, come in four types: Unsubsidized Stafford, Subsidized Stafford, PLUS, and consolidation loans. Unsubsidized, PLUS, and consolidation loans are available to all borrowers attending eligible institutions, while Subsidized loans are available based on a financial needs test. One main difference between Unsubsidized and Subsidized loans is that interest does not accrue for borrowers of Subsidized loans while they are in school. PLUS loans are available to the parents of dependent undergraduate students (Parent PLUS) and to graduate and professional students. Independent undergraduate students are not eligible for PLUS loans, but are allowed to borrow additional Stafford loans up to higher maximums. Consolidation loans allow students to combine all of their federal loans into one loan to simplify payments. Loan limits are set by legislation, and loans can be used only to meet education expenses like tuition and other costs of attendance. Fees for Direct Loans were raised slightly following the 2013 budget sequestration; borrowers are charged an origination fee of 1 percent for Stafford Loans and 4 percent for PLUS loans.

Table 1 provides an overview of these federal borrowing programs. The table presents the total loan balance in each fiscal year (over $1.1 trillion in 2014) and the total number of borrowers (roughly 42.8 million in 2014); it also disaggregates those figures into the shares of loans and borrower types each year. Over time, graduate loans and Parent PLUS loans have increased as a share of federal lending. In 1994, about 68 percent of the portfolio was undergraduate loans, and by 2014 that rate had declined to about 59 percent of the portfolio. In terms of numbers of borrowers, the same growth in graduate and parent loans is apparent. However, much of the increase in graduate debt is held by a rising share of students taking out both graduate and undergraduate loans. The persistence of borrowing at the undergraduate and graduate levels, and the increases in graduate and parent loans (whose loan amounts are limited only by costs of attendance), prove to be important reasons why aggregate and per-student loan amounts increase over time.

Prior to the Federal Direct Loan program, the Federal Family Education Loan (FFEL) program also disbursed federally guaranteed loans through private lenders following lending rules for federally guaranteed loans. The main difference between the programs was financing through

Table 1. Aggregate Loan Volume and Number of Borrowers by Loan Type, 1985–2014

Fiscal year	*Total balance*[a]	*Percent undergraduates*[b]	*Percent graduates*[c]	*Percent parent*[d]	*Total borrowers*[e]	*Percent undergraduates only*[f]	*Percent graduates and undergraduates*[g]	*Percent graduates only*[h]	*Percent parent only*[i]
1985	64,406	76	23	1	8,919	84	6	8	2
1986	72,403	76	23	2	10,193	84	6	8	2
1987	79,209	75	23	2	11,411	85	6	8	2
1988	87,330	75	23	2	12,714	85	5	7	3
1989	95,156	74	23	3	13,938	84	5	7	3
1990	101,450	73	24	3	14,896	84	5	7	4
1991	109,914	72	25	4	15,577	83	5	7	5
1992	118,315	70	25	4	16,035	83	5	7	5
1993	130,637	70	26	5	16,402	82	5	7	6
1994	150,137	68	27	5	17,058	82	5	7	6
1995	172,258	67	27	5	17,585	81	5	7	6
1996	196,827	66	28	6	18,270	81	5	7	7
1997	225,185	65	29	6	18,939	81	6	7	7
1998	258,336	65	29	6	19,934	80	6	7	7
1999	286,876	65	29	6	20,563	80	6	7	7
2000	311,449	64	30	7	21,063	79	7	7	7
2001	340,945	63	30	7	21,661	79	7	7	8
2002	369,617	63	30	7	22,590	78	7	7	8
2003	404,376	62	30	7	23,846	78	8	7	8
2004	441,283	61	31	8	25,285	77	8	7	8
2005	478,881	61	31	8	26,745	77	9	6	9
2006	525,178	60	32	8	28,095	76	9	6	9

(*continued*)

Table 1. Aggregate Loan Volume and Number of Borrowers by Loan Type, 1985–2014 (*Continued*)

Fiscal year	*Total balance*[a]	*Percent undergraduates*[b]	*Percent graduates*[c]	*Percent parent*[d]	*Total borrowers*[e]	*Percent undergraduates only*[f]	*Percent graduates and undergraduates*[g]	*Percent graduates only*[h]	*Percent parent only*[i]
2007	568,054	60	32	8	29,478	76	10	6	9
2008	634,453	59	33	8	31,093	76	10	6	9
2009	706,894	59	33	8	33,480	76	10	6	9
2010	791,637	59	33	8	35,949	76	11	6	9
2011	878,741	59	33	8	38,301	75	11	6	9
2012	966,236	59	33	8	40,174	76	11	6	9
2013	1,050,000	59	33	8	41,529	76	11	6	8
2014	1,130,000	59	34	8	42,760	76	11	6	8

Source: U.S. Treasury tabulations of 4-percent NSLDS sample.
a. Total loan balances in millions of 2014 dollars as of the end of each fiscal year.
b. Percent of total loan balances going toward undergraduate degrees.
c. Percent of total loan balances going toward graduate degrees.
d. Total Parent PLUS loan balances as a percent of total loan balances.
e. Total number of borrowers in thousands as of the end of each fiscal year.
f. Percent of borrowers with only undergraduate balances.
g. Percent of borrowers with both undergraduate and graduate balances.
h. Percent of borrowers with only graduate balances.
i. Percent of borrowers with only Parent PLUS loans.

private capital or direct federal funds, and students saw few differences in lending rules.[9] The Perkins loan program provides additional loans to low-income borrowers with exceptional financial need. Perkins loans make up a small share of federal student loan programs, and the analysis in this paper excludes them.

Interest rates are set by Congress and were identical for Stafford borrowers under the Direct Loan and FFEL programs, but can vary for graduate and undergraduate borrowers. Historically, undergraduate Stafford loan interest rates have been both fixed and variable, depending on the year, and rates have varied between 8.25 percent (1999) and 3.4 percent (2004). In the 2010–11 academic year, interest rates were 6.8 percent; in 2015, they dropped to 4.25 percent.

For most loans, after leaving school, repayment begins after a six-month grace period. Once repayment begins, payment can be stopped through either deferment or forbearance. Loans can go into deferment if a borrower re-enrolls in school, becomes unemployed (for up to 3 years), faces economic hardship, or joins the military or the Peace Corps. Forbearance also allows borrowers to defer loans for up to one year if they are ill, face financial hardship, or perform national service. Interest typically continues to accrue while loans are in deferral or forbearance.[10]

The standard repayment plan for student loans is a 10-year plan. Extended repayment plans of up to 25 years are also available to many borrowers with large balances. In addition, income-based and income-contingent repayment options are available to many borrowers with low incomes and high relative debt burdens. Historically, take-up of income-driven repayment plans has been low, although it has been rising in recent years. Bruce Chapman (1997, 2006) provides a discussion of many of the theoretical issues related to income-contingent repayment plans as well as an overview of income-contingent repayment plans in an international context.

Under the Income-Based Repayment and Pay As You Earn plans, borrowers pay the lesser of 10–15 percent of their income or their payment under a 10-year plan. Under the Income-Contingent Repayment plan, borrowers pay the lesser of 20 percent of their discretionary income or what

9. The rules are almost identical for Direct Loans and FFEL loans, the main difference being that the source of funds is private under the FFEL program. Interest rates for Parent PLUS loans differed slightly in some years, as did eligibility for income-based repayment plans and loan consolidation.

10. The Department of Education provides further information on repayment plans, forbearance, and deferment at https://studentaid.ed.gov/sa/repay-loans/deferment-forbearance.

they would pay under a fixed repayment plan. In many cases, borrowers can pay more under an income-contingent plan than they would under the standard plan, but in some cases loan balances can be discharged after 25 years in repayment. Eligibility for these programs has varied historically depending on the type of loan, time of entry into borrowing, entry into repayment, and debt-to-income ratios. In 2014, the Pay As You Earn plan was made available to all borrowers regardless of entry into borrowing.[11]

Students who miss a payment are considered delinquent, and servicers are required to report delinquency to credit bureaus within 90 days. Loans are in default if delinquent for more than 270 days. Unlike other consumer loans, student loans are nearly impossible to discharge in bankruptcy. If a borrower goes into default, she loses eligibility for deferment, forbearance, and alternative repayment plans, and the loan is assigned to a collection agency. The borrower is then liable for late fees, collection costs, and accruing interest. The government is obligated to collect defaulted loan amounts using wage garnishment of up to 15 percent of the borrower's wages, and through the Treasury Offset Program, which withholds any tax refunds and certain other payments, like a portion of Social Security retirement or disability benefits. Accounting for collection costs, the Department of Education (2014) estimates that recovery rates using these methods were approximately 75 to 85 percent in 2014, compared to recovery rates of 70 percent for mortgage loans (Downs and Xu 2015). Historically, recovery rates have been lower; Deborah Lucas and Damien Moore (2010) estimate recovery rates of around 50 percent in the early 2000s.

II. New Administrative Data Sources

The estimates presented in this paper derive from a random 4-percent sample of federal student loan borrowers assembled from components of the National Student Loan Data System (NSLDS), which is the primary data system used to administer the federal loan programs described above. This data system maintains the information needed to run the loan programs, including the repayment system; assessing eligibility for loans using information from financial aid applications; disbursing loans to institutions based on the students' academic level; tracking when students withdraw or graduate to determine when they must begin repayment and if and when they enter deferment, forbearance, or alternative repayment plans; and

11. The Department of Education provides further information on income-driven repayment plans at https://studentaid.ed.gov/sa/repay-loans/understand/plans/income-driven.

providing the financial accounting of loan balances, interest accrual, transactions, and other changes in loan status. Hundreds of individual pieces of information contained in multiple databases, drawn from hundreds of millions of individual records of aid applications, loan transactions, and status updates are distilled into about 46 million annual observations on 4 million borrowers.

The panel, which follows the same borrowers over time starting from when they first take out a federal student loan, is based on data originally constructed by the Department of Education's Budget Service Division for use in budget projections. These files include information on student characteristics derived from each Free Application for Federal Student Aid (FAFSA) filed by students; information on each loan disbursed by Federal Student Aid (FSA), including the loan balance, its status, and changes in status over time; the institution the loan was disbursed to; and information on Pell Grants received. Information on each borrower obtained from the FAFSA is generally available only for loans originated after fiscal year 1995. However, most of the basic loan information (such as loan amounts and dates of origination, repayment, and default, and institution of study) is available from all sample borrowers starting in fiscal year 1969.[12] The sample is representative of more than 99 percent of federal loans and borrowers. However, while we include Parent PLUS loans in our tabulations of borrowing amounts, when examining the experiences of borrowers as they complete school, enter the labor market, and begin repaying their loans, we focus exclusively on student borrowers and exclude outcomes (and economic status) of parent borrowers.

These data are merged to a panel of administrative earnings and income records that span the (calendar year) period from 1999 to 2014 (data for 2014 are incomplete and preliminary). The primary data of interest are the earnings and total incomes of borrowers. Individual earnings are derived from information reports from employers (W-2s) and from self-employment earnings reported on the Schedule C of individual tax returns. Total income is the sum of all income sources reported by taxpayers; if the taxpayer is married and filing a joint return, this includes any income and earnings of the spouse. In addition, information on filing status and the number of dependent children and federal poverty levels is used to construct indicators of poverty.[13]

12. The sample does not include Perkins loans, which were approximately 1 percent of loans disbursed in 2014 (College Board 2014).

13. Online appendix A provides additional detail on the data, sample, and variable construction.

II.A. Sample Construction

To examine the dynamics of borrowing in the loan market, we focus on the flows of borrowers as they enter borrowing (when they originate their first federal loans) and when they enter repayment (when they start repaying their loans) and look at the relationship of those flows to the overall stock of federal loans and to aggregate student-loan outcomes. We treat the year a loan entered into repayment as a primary focus because that year is typically the first time a student exits school, enters the labor market, is required to make payments, and first becomes liable for delinquency and default. Specifically, we define "entering repayment" as the time when a borrower's last loan enters into repayment, that is, when all of a borrower's loans are in repayment. We define repayment cohorts based on the fiscal year each borrower entered repayment. These definitions closely approximate the aggregate measures of debt, default, and average loan burdens produced by the Department of Education (2014).

We define entrants as first-time borrowers, assign them to entry cohorts based on the fiscal year their first loans were originated, and use information on the students from the first loan-related FAFSA filed and the institution that originated the loan. This provides a consistent measure of new originations and borrower characteristics when borrowers first enter the loan system. Defining flows based on first-time and last-time borrowing obscures the fact that an educational career sometimes involves multiple spells of borrowing (re-entry) as students take time off, change institutions, or go to graduate school. In practice, however, this convention has little effect on our analysis of borrowers, because loan outcomes are strongly correlated within borrowers (that is, when borrowers default they default on all loans) and because most variation in loan outcomes occurs after the borrower leaves school and permanently enters the labor force. In addition, the fact that spells of borrowing may overlap at the end of the sample period introduces censoring effects. For example, some borrowers entering repayment in 2014 will subsequently return to graduate school.

II.B. Variable Construction

Most variables used in our analysis are straightforward; characteristics of borrowers, like family income, age, and gender, are taken directly from the FAFSA. Data on the neighborhoods of borrowers—local unemployment rates, poverty rates, median household income, and percent black, white, and Hispanic—are derived by matching the ZIP code provided on the first FAFSA with ZIP code–level statistics from the 2000 Decennial Census. Loan information, such as disbursements and balances, are the

sum of all Direct and FFEL undergraduate and graduate loans or Parent PLUS loans at the end of the fiscal year for each borrower. We use these loan types and the reported academic level of borrowers to differentiate undergraduate from graduate borrowers. All dollar amounts are in real 2014 dollars (adjusted using the personal consumption expenditure deflator) unless noted otherwise.

In practice, over the course of an educational career students may attend multiple institutions and take out both undergraduate and graduate loans. Our general approach is to classify students based on their characteristics, the institution they attended, and their level of school when they borrow for the first time. Hence, borrowers may be classified as attending a 2-year school even though some may ultimately complete a 4-year degree elsewhere, and borrowers who start their education (and student loan borrowing) at 4-year institutions may ultimately go on to graduate or professional schools. One implication of this choice is that when we examine the eventual loan burden of a student starting to repay her loans, her loan burden may include a combination of undergraduate and graduate loans even though she started off as an undergraduate borrower.

In practice, this assumption has little effect on our conclusions because changes in enrollment between sectors are relatively rare. For example, a borrower attending a 4-year institution is generally likely to complete his education there. As a result, there is little difference in outcome measures like default rates by institution type whether they are based on first institution or last institution attended.

However, for certain loan measures, such as the number of borrowers or the amounts borrowed by institution type, whether to measure based on first or last institution attended has a greater effect, particularly for nontraditional borrowers. In particular, some individuals who started at 2-year institutions went on to attend 4-year institutions or for-profit institutions where they accumulated larger debts, and some students who started as traditional borrowers later returned to school and last attended a for-profit school, especially during the recession. Qualitatively, the enrollment and borrowing patterns are quite similar whether we use first or last institution attended, but the levels can be somewhat different.[14]

One important advantage of these data over other sources is the availability of information on the institutions that students borrowed to attend, including the specific schools, the controls (public, private, or for-profit),

14. The online data appendix provides complete tabulations of borrower outcomes and debts using both first and last institution attended.

and types (2-year or 4-year). To illustrate the role of institutions, we present much of our analysis based on the type of institution attended. In particular, we use a common index of selectivity from Barron's Educational Series (2008) to segment students based on the control, type, and selectivity of the institutions they attended, and their level of study into six broad groups: for-profit institutions; 2-year public and private institutions (the vast majority of which are community colleges); nonselective 4-year public and private institutions (schools that Barron's reports as admitting more than 85 percent of applicants); somewhat selective institutions (which admit 75 to 85 percent of applicants); selective institutions (which admit fewer than 75 percent of applicants); and graduate-only borrowers (borrowers whose first and only loans were graduate loans).[15]

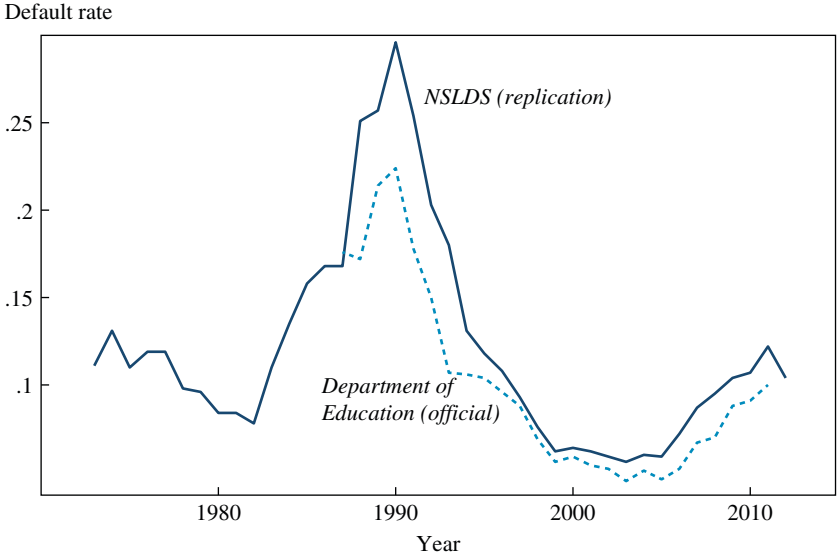
Our primary indicator of student loan distress is the 3-year cohort default rate measured as the fraction of borrowers entering repayment in a fiscal year who are in default on a federal loan within 3 years (1,095 days) from the date the loan entered repayment. We introduce several other indicators of student loan burdens or delinquency, including debt-service-to-earnings ratios and rates of negative amortization, which we define as the fraction of student loan borrowers who owe more on their loans at a specified time after they entered repayment.

These estimates of the aggregate loan volume, number of borrowers, and the cohort default rate closely mirror the official measures produced by the Department of Education. Figure 1 shows default rates over time in our sample, compared with aggregate statistics released by the Department of Education.[16] The replicated default rate closely matches the pattern of published statistics for most of the overlapping period. However, our replication clearly differs slightly for several reasons that relate to our sample construction and to the construction of the official default rate. First, our sample is a person-by-year sample, which means any student appears only once. In official statistics, a borrower who attends multiple institutions may be included multiple times. Second, we focus on the last time a borrower enters repayment, when default risks are likely to be higher. Finally, our sample includes all institutions, all undergraduate and graduate loans, and all (nonparent) borrowers, which is a broader array of programs and

15. Our nonselective group corresponds to Barron's "Non-Competitive or Less-Competitive"; somewhat selective to its "Competitive"; and selective to its "Very, Highly, and Most Competitive." The online data appendix also provides estimates based on the first and last type and control of the institution attended.

16. We use 2-year cohort default rates for this validation exercise since historical data on this series exist covering a longer period of time.

Figure 1. Two-Year Cohort Default Rate, 1972–2012[a]



a. This figure compares the official 2-year cohort default rate from the Department of Education for 1992–2010 (dashed) with our replication based on the 4-percent NSLDS sample from 1972–2012 (solid). In our replication, the 2-year cohort default rate is the fraction of federal student loan borrowers entering repayment on all federal loans who default by the end of the next fiscal year after the fiscal year that they entered repayment. Cohorts are defined by fiscal year entered repayment.

students than is counted in the official rate. For instance, some students may be excluded from the official rate if their institution demonstrated they met certain criteria for exclusion. Prior to 1995, direct loans were excluded from the official default rate, and borrowers from recently closed institutions appear to have been excluded, which may have affected estimates during periods when the number of participating institutions was declining sharply. For the aforementioned reasons, our estimates are likely to be somewhat higher than the official rate, particularly in earlier years.

III. The Rise of Nontraditional Borrowing and Its Consequences

A primary focus of our analysis is on the divergent outcomes of what we call nontraditional and traditional borrowers. In this section, we describe how nontraditional borrowing increased in recent years, compare this increase in borrowing to increases in enrollment measured in other data sources, examine how the increase in nontraditional borrowing affected the

composition of the federal loan portfolio, and illustrate how these changes in borrowing patterns over the recession resulted in a sharp increase in the number of nontraditional borrowers out of school and into the labor market as the recession waned.

III.A. New Borrowers at For-Profit Schools, 2-Year Institutions, and Other Nonselective Institutions

As discussed earlier, we define traditional borrowers as those attending 4-year public and private institutions because they represent one case of a "typical" college student: They start college in their late teens, soon after completing high school, are dependent on their parents (or assumed to be) for aid purposes, pursue 4-year degrees and, frequently, head on to graduate study. The median age of first-time undergraduate borrowers at these schools is 19 (26 for graduate borrowers) and more than 80 percent of undergraduates at relatively selective institutions are dependents for purposes of financial aid.

In addition, these borrowers are traditional in that historically they represented a large share of federal borrowers and loan amounts. In 1999, borrowers at 4-year public and private institutions and graduate-only borrowers represented about 70 percent of new borrowers, about two-thirds of all federal student loan borrowers (the stock), and almost 80 percent of aggregate student loans outstanding. One reason for the outsized influence of 4-year public and private institutions is that these institutions, particularly the most selective private institutions and graduate professional schools, were relatively more expensive, hence students there had a greater need to borrow to attend.

Nontraditional borrowers, in contrast, constituted only a small share of federal student loan borrowers and an even smaller share of the aggregate student loan portfolio. For instance, full-time undergraduate students at for-profit schools were a relatively small share of all new full-time students in 2000 (10 percent) and an even smaller share of total full-time plus part-time enrollment (3 percent). While almost all for-profit students took out federal loans to study, their relatively small share of enrollment meant that they still represented only 20 percent of new borrowers, 14 percent of active borrowers, and 13 percent of outstanding federal loans. Hence, they were a relatively small share of both students and student-loan dollars, though a disproportionate (but still modest) share of borrowers.

While 2-year borrowers—primarily community college students—were also a small share of federal borrowers and loan amounts, the reason was quite different. Community college students rarely borrowed and when

they did, they borrowed relatively small amounts. In 2000, community college students were 27 percent of new fall enrollment and 43 percent of total postsecondary undergraduate enrollment—more than double the enrollment share of private nonprofit institutions and almost 15 times the reported enrollment of the for-profit sector. However, according to the Department of Education, in the 2000–01 school year, only 15 percent of new 2-year public students borrowed.[17] As a result, they accounted for only 9 percent of active undergraduate federal borrowers in 2000 and, because their average loan amounts were relatively small, they accounted for only about 4 percent of undergraduate federal loan originations that year and 8 percent of all outstanding federal loans.

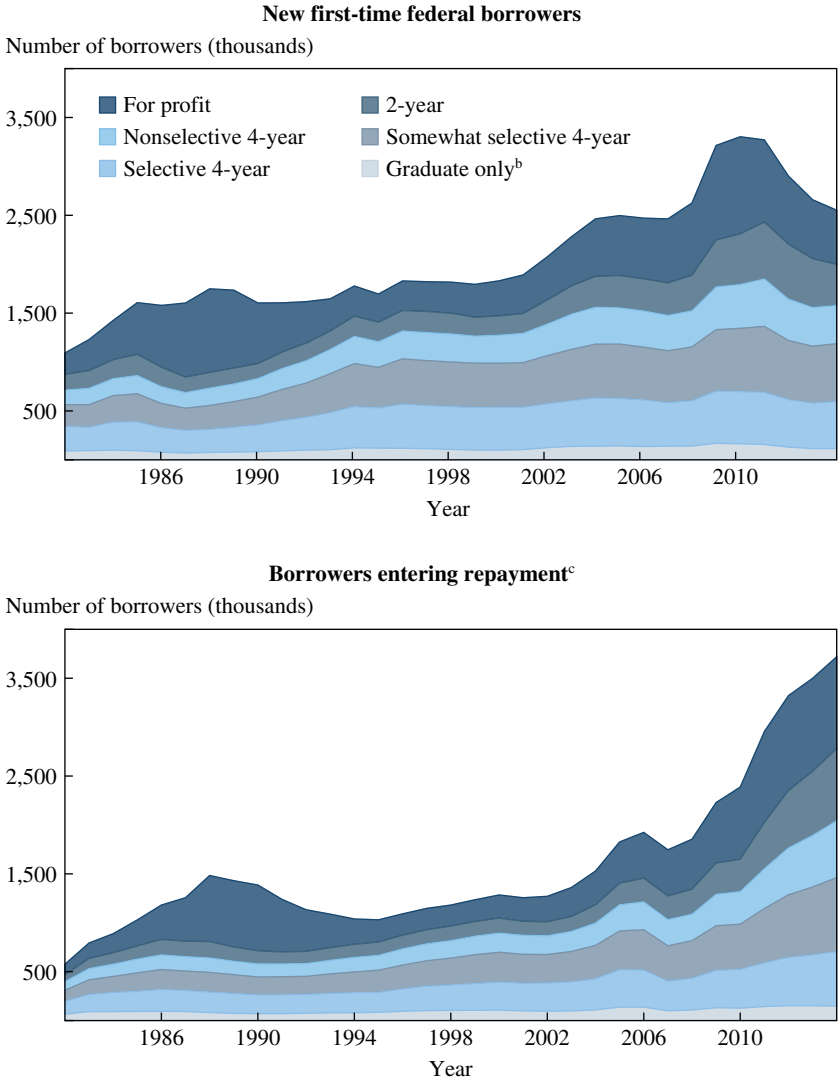
More generally, nontraditional borrowers are more likely to be older at entry; their median age at first borrowing is 24 at for-profit schools and 23 at 2-year institutions. They are more likely to be independent for financial aid purposes, which means they are subject to higher borrowing limits and less likely to draw on the support of their parents. While the characteristics of the students themselves are therefore relatively similar at for-profit and 2-year schools, the share that leave with loans and the average loan burden is much lower among community college students.

Of course, categories based solely on an institution's ownership or control and the predominant type of degree awarded miss some heterogeneity within and across groups. Nonselective 4-year public and private institutions are a particularly heterogeneous group and include many institutions whose students have background characteristics, borrowing rates, and loan and labor market outcomes more similar to students at for-profit institutions than to those at more selective 4-year institutions. They may also be unconventional in other dimensions, such as having a predominate focus on online education. The estimates we provide of the outcomes of students from nonselective 4-year public and private institutions therefore reflect an average of potentially disparate outcomes within that borrowing population.

To illustrate the changes in borrowing over time by type of institution, figure 2 (top panel) provides more perspective on the rise (and decline) of borrowing using estimates of the number of first-time borrowers at each type of institution each fiscal year. This figure shows the steady growth of the for-profit sector over the last 15 years and, especially, the surge in enrollment during the recession. In 2009 and 2010, borrowers at for-profit

17. See the National Center for Education Statistics' online "Digest of Education Statistics," table 331.20 (https://nces.ed.gov/programs/digest/d14/tables/dt14_331.20.asp).

Figure 2. First-Time Student Borrowers and Borrowers Entering Repayment by Institution Type, 1982–2014[a]



Source: U.S. Treasury tabulations of 4-percent NSLDS sample.

a. For both panels, institution type and selectivity (as defined by Barron's Educational Series) refer to institution borrower first attended. All types, except for-profit, include both public and private institutions. See text for details.

b. Borrowers taking out only graduate loans.

c. Borrowers entering repayment on their last federal loan; that is, all loans have entered repayment.

schools represented roughly 30 percent of new borrowers (about 980,000 new borrowers each year), and new borrowers at 2-year schools represented roughly 16 percent of new borrowers (and 19 percent at its peak in 2012 of 580,000 borrowers). Between 2006 and their respective peaks, the number of new borrowers at 2-year schools jumped by 71 percent, and the number at for-profit schools by 60 percent. While new borrowing also increased at public and private 4-year institutions during the recession, the increases from 2006 to their relative peaks were much smaller: 31 percent at nonselective schools (2011 peak), 12 percent at the most selective schools (2011), and 24 percent among graduate-only borrowers (2009). From 2009 to 2011, almost half of all new federal borrowers (45 percent) were students at either for-profit or 2-year schools.

III.B. Comparison of Enrollment and Borrowing in NCES and NSLDS

The increase in and importance of nontraditional borrowers is less visible in official statistics that report the level of total or new enrollment. Many nontraditional borrowers appear to enroll in part-time or certificate programs or enroll outside the usual academic calendar, which starts in the autumn, and therefore they appear to fall outside the definitions used in official enrollment statistics. In addition, because borrowers at 2-year institutions and for-profit institutions enroll for much shorter durations than those at other 4-year institutions, the annual level of enrollment undercounts the number of individual borrowers flowing through for-profit and community colleges.

Table 2 provides a basic comparison of the enrollment data from the National Center for Education Statistics (NCES), which is the primary public source for measuring college enrollment, with NSLDS data. The table compares the NCES measures of total fall undergraduate enrollment (including full-time and part-time students) to estimates of the number of active undergraduate borrowers and the amount they borrowed from NSLDS in 2000 and 2011. The measures are disaggregated by school type available in the NCES: public 2-year, public 4-year, private nonprofit, and for-profit. The first two columns show total fall enrollment at degree-granting institutions, followed by the number of total active borrowers taking out loans that year. The last two columns show federal undergraduate loan originations.

As table 2 shows, total fall undergraduate enrollment rose from 13.2 million in 2000 to 18.1 million in 2011, a 37 percent increase. During the same interval enrollment at for-profit institutions increased by more than 300 percent, and as a share of all students enrollment at for-profits increased from

Table 2. Undergraduate Enrollment and Borrowing, 2000 and 2011[a]

	Total fall enrollment (millions)[b]		*Total active borrowers (millions)*[c]		*Active originations (billions of 2014 dollars)*[d]	
	2000	*2011*	*2000*	*2011*	*2000*	*2011*
Students, borrowers, and loan amounts						
Public 2-year	5.7	7.1	0.4	1.4	1.6	8.7
Public 4-year	4.8	6.6	2.0	3.3	12.2	24.2
Private nonprofit	2.2	2.7	1.1	1.7	7.6	13.6
For-profit	0.4	1.7	0.6	2.2	3.6	18.0
Total	13.2	18.1	4.0	8.5	24.9	64.5
Shares of students, borrowers, and loan amounts[e]						
Public 2-year	43	39	9	17	4	8
Public 4-year	37	37	49	38	44	38
Private nonprofit	17	15	28	20	40	30
For-profit	3	9	14	25	12	24

Borrowers per student, originations per student, and originations per borrower[f]

	Borrowers per student (percent)		*Originations per student (dollars)*		*Originations per borrower (dollars)*	
Public 2-year	7	20	277	1,226	4,194	6,121
Public 4-year	41	49	2,519	3,649	6,173	7,436
Private nonprofit	52	63	3,431	5,000	6,641	7,998
For-profit	138	131	8,935	10,861	6,476	8,313
Total	31	47	1,894	3,570	6,163	7,553

Sources: Full-time enrollment data come from the U.S. Department of Education, National Center for Education Statistics (NCES), Higher Education General Information Survey, tables 331.20 and 303.70. Borrower and origination data come from the U.S. Treasury tabulations of the 4-percent NSLDS sample.

a. Degree-granting institutions include those that grant associate's degrees or higher, and participate in Title IV federal financial aid programs.

b. Columns 1 and 2 of the top panel show total fall undergraduate enrollment (millions; full- and part-time) in degree programs that participate in Title IV programs. Data come from NCES.

c. Columns 3 and 4 of the top panel show total active undergraduate borrowers (millions) receiving undergraduate loan disbursements in each fiscal year. Data come from NSLDS tabulations.

d. Columns 5 and 6 of the top panel show aggregate federal undergraduate loan originations in billions of 2014 dollars. Data come from NSLDS tabulations.

e. The middle panel shows institution shares (as percentages) of the totals from the top panel.

f. Originations include all loans disbursed on behalf of undergraduate students (excluding PLUS loans).

3 to 9 percent.[18] Enrollment at 2-year public schools increased roughly 25 percent, though as a share of all students it declined from 43 to 39 percent. Prior to and during the recession, many new students enrolled in school, but those students disproportionately enrolled at for-profit institutions and at lower-than-average rates at community colleges and private nonprofit schools. These enrollment patterns contributed to increases in the number of borrowers, not only because they increased the number of eligible students but also because the composition of students shifted toward for-profit schools (where the ratio of borrowers to NCES-reported students is 1.4 to 1) and away from 2-year and private nonprofits, where fewer students tended to borrow.

In addition, the rate of borrowing among NCES-reported enrolled students increased, particularly at 2-year public institutions where borrowing rose from 7 percent of students to 20 percent. Rates of borrowing per student increased by about 20 percent at 4-year public and private institutions (from 41 to 49 percent of students at 4-year public institutions and from 52 to 63 percent at private institutions), and declined slightly at for-profits. Because enrollment at for-profits increased as well as the borrowing rates at 2-year public schools, students at these two types of schools increased as a share of all active borrowers, rising from 9 to 17 percent and from 14 to 25 percent, respectively, between 2000 and 2011. Other things equal, had enrollment increased more evenly across institution types—for instance, had more students attended 2-year schools rather than for-profit schools—and had the increase in the share of community college students who borrowed been smaller, then the number and distribution of student loan borrowers would have been quite different, with a larger share of 4-year public and private borrowers and fewer nontraditional borrowers.

The fact that there are almost 40 percent more for-profit *borrowers* (from NSLDS data) than reported for-profit *students* (from NCES data) is one indication that the NCES enrollment figures understate the number of nontraditional borrowers (see the first two columns of the lower panel of table 2). Several factors are likely to explain this discrepancy, including borrowers in nondegree programs and borrowers enrolled outside the traditional academic cycle. Increases in such students may also be one reason why the number of active borrowers rose relative to NCES measures of

18. This is an underestimate, since the NCES data only include degree-granting institutions, defined as those that grant associate's degrees or higher and participate in Title IV federal financial aid programs. Some for-profits grant only certificates.

enrolled students at 2-year public institutions. Another reason why enrollment statistics understate the number and growth of nontraditional borrowers is that the duration of enrollment differs across institution types. A short-duration program serves more students over a given period of time because its student body turns over more quickly. As a result, for a given level of enrollment at 2-year and for-profit schools, there were disproportionately more new borrowers being produced.[19] Hence, increases in the level of enrollment at for-profit institutions and in borrowing rates at 2-year public schools had outsized impacts on the number and composition of student loan borrowers.

III.C. Implications for the Stock of Borrowers and Debt

The inflow of these new borrowers caused both the stock of outstanding debt and the borrowers to increase in size and change in composition. Table 3 shows the number of outstanding student loan borrowers from 1985 to 2014 by the type of institution they first attended. In 2000, borrowers from for-profit and 2-year institutions accounted for less than 35 percent of all borrowers; by 2014, the number of borrowers had more than doubled at for-profit schools and 2-year institutions, rising by 114 and 167 percent, respectively. The number of borrowers at selective public and private institutions increased by much less, about 69 percent. Almost 45 percent of the increase in the number of borrowers from 2000 to 2014 (when the number of borrowers increased by roughly 21.7 million) were borrowers who started at for-profit institutions (26 percent) and 2-year institutions (18 percent).[20] By 2014, almost 40 percent of all federal borrowers were nontraditional borrowers. Had the number of nontraditional borrowers instead increased at the same rate as the number of traditional borrowers over this period, there would have been roughly 3.5 million (21 percent) fewer nontraditional borrowers. Moreover, the data on enrollment rates and borrowing-per-student in table 2 suggest that if a greater share of rising enrollment had occurred at 2-year public institutions or at 4-year public and

19. For example, compare a scenario in which there are two million students and one million new borrowers each year to a scenario in which there are two million students and two million new borrowers each year. In the latter case, each year there must be twice as many students exiting into the labor market with student loans.

20. Based on the institution a borrower *last* attended, the contribution of nontraditional institutions is somewhat larger, with for-profit and 2-year schools representing 29 and 15 percent of the increase, respectively. This suggests that there is substantial overlap between the for-profit and 2-year nonprofit sectors, with many students beginning borrowing spells at 2-year nonprofits and later borrowing at for-profit institutions.

Table 3. Number of Federal Borrowers by Institution Type, 1985–2014[a]

Fiscal year	*Total*	*For-profit*	*2-year*	*Non-selective 4-year*	*Somewhat selective 4-year*	*Selective 4-year*	*Graduate only*[b]
1985	8,919	1,764	1,072	1,309	1,827	2,182	765
1986	10,193	2,322	1,232	1,434	2,011	2,371	824
1987	11,411	2,983	1,340	1,533	2,163	2,523	869
1988	12,715	3,721	1,439	1,649	2,317	2,668	918
1989	13,938	4,374	1,533	1,759	2,485	2,821	962
1990	14,896	4,815	1,603	1,858	2,645	2,971	1,000
1991	15,578	5,055	1,661	1,948	2,801	3,087	1,022
1992	16,036	5,153	1,700	2,029	2,936	3,185	1,028
1993	16,403	5,102	1,740	2,118	3,107	3,305	1,026
1994	17,060	5,100	1,820	2,242	3,343	3,492	1,058
1995	17,591	5,056	1,886	2,351	3,537	3,661	1,093
1996	18,278	5,027	1,965	2,481	3,786	3,880	1,126
1997	18,949	4,997	2,052	2,620	4,025	4,084	1,156
1998	19,946	5,054	2,174	2,805	4,332	4,351	1,214
1999	20,572	5,016	2,237	2,942	4,572	4,556	1,233
2000	21,073	4,995	2,281	3,062	4,762	4,714	1,242
2001	21,673	5,043	2,345	3,201	4,958	4,858	1,251
2002	22,604	5,180	2,469	3,378	5,213	5,052	1,293
2003	23,857	5,426	2,649	3,606	5,521	5,279	1,358
2004	25,297	5,772	2,857	3,845	5,844	5,537	1,422
2005	26,756	6,140	3,076	4,075	6,172	5,782	1,491
2006	28,108	6,506	3,279	4,292	6,460	6,000	1,548
2007	29,490	6,907	3,491	4,502	6,755	6,196	1,617
2008	31,106	7,408	3,738	4,730	7,085	6,430	1,692
2009	33,491	8,181	4,123	5,055	7,533	6,776	1,802
2010	35,960	8,984	4,551	5,391	8,004	7,110	1,900
2011	38,312	9,617	5,034	5,756	8,483	7,419	1,982
2012	40,200	10,119	5,487	6,049	8,865	7,637	2,023
2013	41,555	10,425	5,834	6,278	9,170	7,795	2,033
2014	42,792	10,704	6,101	6,497	9,479	7,951	2,041

Source: U.S. Treasury tabulations of 4-percent NSLDS sample. Selectivity data come from Barron's Educational Series (see text).

a. Total borrowers (in thousands) with outstanding federal loan balances each fiscal year, by institution type of first borrowing.

b. Refers to borrowers who started borrowing at the graduate level.

private institutions, where borrowing rates are relatively low, rather than at for-profits, or had borrowing among community college students increased at rates closer to the rates at 4-year public institutions, then there would have been even fewer nontraditional borrowers.

Similarly, the increase in the number of nontraditional borrowers is an important contributor to the increase in overall debt and to the share

Table 4. Aggregate Federal Student Loan Debt by Institution Type, 1985–2014[a]

Fiscal year	*Total*	*For-profit*	*2-year*	*Non-selective 4-year*	*Somewhat selective 4-year*	*Selective 4-year*	*Graduate only*[b]
1985	64,406	8,151	5,721	10,005	12,832	18,199	9,496
1986	72,403	10,598	6,587	10,791	14,230	19,950	10,247
1987	79,209	13,507	7,059	11,234	15,223	21,280	10,889
1988	87,330	17,372	7,529	11,808	16,251	22,588	11,731
1989	95,156	20,853	8,003	12,407	17,375	23,932	12,508
1990	101,450	22,762	8,339	12,944	18,618	25,389	13,320
1991	109,914	24,334	8,937	13,976	20,572	27,557	14,460
1992	118,315	25,225	9,568	15,141	22,744	29,903	15,655
1993	130,637	26,254	10,450	16,894	26,011	33,545	17,403
1994	150,137	27,214	11,773	19,666	31,166	39,666	20,564
1995	172,258	28,223	13,290	22,866	36,964	46,649	24,156
1996	196,827	29,608	14,953	26,437	43,638	54,282	27,763
1997	225,185	31,419	17,072	30,652	51,322	63,023	31,541
1998	258,336	35,641	19,874	35,916	59,600	71,737	35,398
1999	286,876	38,412	21,912	40,391	67,253	79,988	38,749
2000	311,449	40,359	23,711	44,467	74,248	87,225	41,261
2001	340,945	43,586	26,029	49,472	82,342	95,257	44,076
2002	369,617	47,330	28,397	54,257	89,998	102,775	46,655
2003	404,376	51,961	31,520	60,029	99,066	111,694	49,885
2004	441,283	57,519	34,805	66,042	108,275	120,905	53,495
2005	478,881	62,973	38,369	71,946	117,487	129,882	57,928
2006	525,178	70,222	42,816	79,351	128,726	140,817	62,916
2007	568,054	77,369	47,199	86,024	139,092	150,266	67,765
2008	634,453	89,058	54,006	96,248	154,850	165,061	74,904
2009	706,895	104,206	62,109	107,278	171,479	179,847	81,657
2010	791,638	122,070	72,125	119,963	190,672	197,040	89,452
2011	878,741	141,322	83,431	133,518	210,170	213,308	96,683
2012	966,237	159,022	95,208	147,642	230,609	229,943	103,502
2013	1,045,180	174,097	106,293	160,835	249,924	244,667	109,068
2014	1,125,652	188,858	117,320	174,504	270,309	259,666	114,709

Source: U.S. Treasury tabulations of 4-percent NSLDS sample. Selectivity data come from Barron's Educational Series (see text).

a. Total outstanding federal loan balances in millions of 2014 dollars for borrowers each fiscal year, by institution type of first borrowing.

b. Refers to borrowers who started borrowing at the graduate level.

of debt owed by students from for-profit and 2-year institutions. Table 4 shows aggregate federal student loan debt by institution type first attended. Between 2000 and 2014, the amount of debt owed by borrowers who first attended a for-profit institution more than quadrupled from $40 billion to $189 billion, and nearly quintupled from $24 billion to $117 billion among borrowers who had first attended a 2-year public institution. The

share of outstanding loan balances attributable to for-profit school students increased from 13 to 17 percent over that period, and from 8 to 10 percent among 2-year college students.[21] Loan volumes also increased rapidly at more selective institutions, although the rate of increase was slower. The share of loans owed by borrowers from 4-year public and private institutions and graduate-only borrowers correspondingly fell by almost 7 percentage points from 79 to 73 percent. Had the number of nontraditional borrowers merely increased at the same rate as traditional borrowers, and had per-student borrowing increased by the same amount, then total debt owed by nontraditional borrowers would have been $94 billion (31 percent) lower.

A more concrete picture of how changes in nontraditional borrowing have shaped the student loan market is evident in table 5, which presents estimates of the cumulative debts of students according to the institutions they last attended in 2000 and 2014 for the 25 institutions whose students owed (collectively) the most in federal student loan debt. The debt of the students associated with each institution in each of the years includes the cumulative federal loan liabilities of borrowers, including undergraduate, graduate, and parent loans, plus any accrued interest. In addition, the analysis assigns each borrower's cumulative debts—including any debts incurred at previous institutions—to the last institution of borrowing.[22] In 2000, with the exception of the University of Phoenix, all the institutions were either 4-year public or private nonprofit institutions, often state flagship universities and institutions with large graduate programs. In 2014, 8 of the 10 institutions whose students owed the most aggregate debt and 13 of the top 25 were for-profit institutions, and one

21. Based on the institution *last* attended, the increase is from 12 to 20 percent for for-profit institutions and 4 to 6 percent at 2-year institutions.

22. This method of assigning loan debts provides a useful measure of how the characteristics of the loan portfolio changed as a result of shifts in enrollment, because many students borrow to attend only one institution and because students' labor market outcomes and loan performance are closely related to the institution last attended. As a practical matter, this method of assignment avoids the challenges of tracing accrued interest or subsequent payments to individual institutions. However, this method overstates the amount of debt originated at certain institutions, especially those with a disproportionate number of graduate student borrowers or older borrowers, and it understates debts originated at certain primarily undergraduate institutions. The online data appendix provides an alternative measure of total cumulative federal loan disbursements and outcomes by institution and type of loan (excluding interest and payments), which shows a similar pattern of change to that presented in table 5.

Table 5. Cumulative Debt of Student Loan Borrowers by Institution Last Attended, 2000 and 2014

	Institution	*Total debt*[a]	*Total borrowers*	*Institution*	*Total debt*[a]	*Total borrowers*	*5-year cohort, default rate*[b]	*5-year cohort, percent balance repaid*[c]
Rank	*2000*			*2014*				
1	New York University	2,184,601	72,650	University of Phoenix, Phoenix campus	35,529,283	1,191,550	45	1
2	University of Phoenix, Phoenix campus	2,099,828	103,475	Walden University	9,833,470	120,275	7	0
3	Nova Southeastern University	1,736,919	34,900	Nova Southeastern University	8,748,887	94,350	6	–3
4	Pennsylvania State University	1,710,951	123,800	DeVry University, Downers Grove campus	8,249,788	274,150	43	–4
5	University of Southern California	1,609,511	51,525	Capella University	8,043,635	104,450	19	–5
6	Ohio State University, Columbus campus	1,533,954	82,250	Strayer University, Washington campus	6,693,570	144,400	31	–6
7	Temple University	1,531,762	59,900	Kaplan University, Davenport campus	6,664,067	220,125	53	0
8	Arizona State University, Tempe campus	1,385,858	70,675	New York University	6,307,264	110,775	6	34
9	Michigan State University	1,321,997	65,650	Argosy University, Chicago campus	6,179,207	104,325	15	–7
10	University of Minnesota, Twin Cities campus	1,289,873	66,675	Ashford University	5,891,799	205,000	47	2
11	Boston University	1,289,257	50,850	Grand Canyon University	5,881,420	145,850	36	0
12	University of Texas, Austin campus	1,264,226	64,650	Liberty University	5,678,555	142,875	14	14

13	University of Florida	1,186,645	52,050	University of Southern California	5,340,123	83,400	5	20
14	University of California, Los Angeles campus	1,159,430	54,975	Pennsylvania State University	5,310,636	210,125	14	21
15	University of Michigan, Ann Arbor campus	1,126,159	44,725	Arizona State University, Tempe campus	4,928,019	158,800	17	12
16	Columbia University	1,120,001	31,225	ITT Technical Institute	4,618,538	191,225	51	–1
17	University of Pittsburgh, Pittsburgh campus	1,106,448	48,925	Ohio State University, Columbia campus	4,362,143	132,725	12	19
18	Indiana University, Bloomington campus	1,101,234	53,225	Temple University	4,251,334	100,500	12	13
19	Rutgers University, New Brunswick campus	1,077,418	60,150	DeVry University, Keller Graduate School of Management	3,900,283	49,375	13	1
20	University of Pennsylvania	1,033,615	33,300	American InterContinental University, online campus	3,735,319	129,850	41	–3
21	University of Arizona	983,809	45,975	University of Minnesota, Twin Cities campus	3,679,264	101,650	7	18
22	University of Wisconsin, Madison campus	981,553	45,050	Michigan State University	3,596,661	99,925	11	14
23	Florida State University	976,114	49,125	Rutgers University, New Brunswick campus	3,436,474	116,925	9	19
24	Virginia Commonwealth University	965,668	39,425	Colorado Technical University, Colorado Springs campus	3,300,070	114,000	47	1
25	University of Washington, Seattle campus	954,589	51,625	Indiana University–Purdue University Indianapolis	3,141,584	74,500	15	10

Source: U.S. Treasury tabulations of 4-percent NSLDS sample.

a. Total volume of student loans outstanding in thousands of 2014 dollars.

b. Percentage of the 2009 repayment cohort that had defaulted by 2014.

c. Percentage of the total balance for the 2009 repayment cohort that had been repaid by 2014. Negative numbers indicate balance has increased.

private nonprofit institution was largely an online program.[23] At certain institutions, a majority of the debt was accumulated by graduate and professional students with high average balances, such as at Walden University, Nova Southeastern University, New York University, and the University of Southern California (at these institutions a larger share of the total debt owed by students was accumulated at prior institutions). At other institutions, almost all of the debt was undergraduate debt, such as at the University of Phoenix, Strayer University, Kaplan University, Ashford University, and ITT Technical Institute.

The final two columns show, respectively, the fraction of students that defaulted and the fraction of their initial balances repaid by 2014 for the 2009 cohort. First, there is substantial heterogeneity across institution types in terms of default rates and balances repaid, with students at some for-profits experiencing 5-year default rates approaching 50 percent. Second, while the share of balances repaid and default rates are highly correlated, students from some schools maintain low default rates despite not paying down their debts. This could be due to deferment, forbearance, the use of income-based repayment plans, or other plans that allow borrowers to suspend or reduce their payments without risk of default.

III.D. Post-Recession Exodus: Rapid Increase in Repayment Flows

The lower panel of figure 2 shows the subsequent flow of borrowers into repayment by institution type. The large increase in borrowing at the onset of the recession subsequently turned into a mass exodus of borrowers into repayment as the recession waned. The exodus was magnified not just by the enrollment patterns during the recession—the fact that many new borrowers sheltered from the labor market by enrolling—but by the durations of their enrollment. Many borrowers from 4-year schools

23. Loan balances from all federal sources (undergraduate, graduate, and parent) are aggregated for each institution with an individual agreement to participate in Title IV programs (specifically by the 6-digit Office of Postsecondary Education ID). For some institutions, this aggregates over many branches or campuses, but some institutions that are a part of larger umbrella organizations are identified separately. Using an alternative method of aggregating related institutions would lead to different but qualitatively similar results. For instance, if aggregated by parent company (for large for-profit chains) or by state university systems (for 2-year or 4-year public institutions), 10 of the 25 educational systems whose students owed the most in aggregate student loan debt in 2014 would be for-profit systems; one is a nonprofit organization (Nova Southeastern University), and the other 14 are state university systems (including Florida, New York, California, Georgia, and Texas).

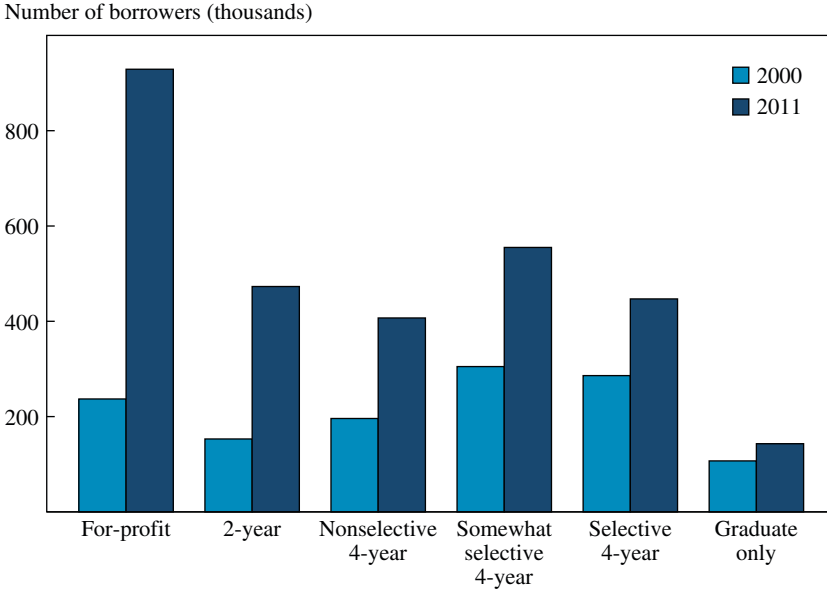
extended their enrollment by staying in school somewhat longer, delaying entry into repayment. The surge in borrowing at the start of the recession was driven by relatively short-duration programs (such as 1-year and 2-year certificates and degrees) or short-duration enrollments as students rapidly dropped out. The confluence of these factors resulted in a spike of borrowers into repayment.

We will return to these dynamics later, but one observation worth noting here is that over the last several years of the period there were simply many more borrowers entering into repayment relative to the number of new borrowers. About 63 percent of borrowers and about 62 percent of the aggregate value of loans were in repayment in 2013. This was up from 53 percent of borrowers and 60 percent of the value of loans in 2007. Hence, a much larger share of borrowers (who hold a much larger share of total loan dollars) are feeling the burden of paying their loans today than in previous years. Moreover, many more borrowers are in the earliest years of repayment, a time when the loan burdens (relative to earnings) are highest and when default rates peak. Even absent any other changes in the loan market, this increase in the number and share of new entrants should be expected to result in high absolute numbers of students in default or struggling in their first years.

In addition to the sheer volume of borrowers entering repayment, the composition of borrowers and the institutions they attended changed substantially. To illustrate the magnitude of these changes, figure 3 compares the number of borrowers entering repayment in 2000 and 2011.[24] The number of (undergraduate and graduate) borrowers entering repayment from for-profit institutions increased rapidly, from about 237,000 in 2000 (18 percent of borrowers) to 930,000 in 2011 (31 percent of borrowers). At 2-year institutions, the number of borrowers entering repayment increased from about 150,000 in 2000 (12 percent) to about 470,000 in 2011 (16 percent). Hence, in 2011 borrowers from for-profit and 2-year schools represented roughly 47 percent of federal student loan borrowers entering repayment. After 2011, the number of for-profit borrowers entering repayment remained above 900,000 through 2014 and the number of 2-year borrowers continued to rise, hitting about 740,000 in 2014.

24. We choose 2011, in particular, for this comparison because it is the last cohort for which we observe both labor market outcomes and 3-year cohort default rates, and because we use this comparison in our decomposition analysis. Comparisons using alternative base years provide qualitatively similar results.

Figure 3. Student Borrowers Entering Repayment by Institution Type, 2000 and 2011 Cohorts[a]



Source: U.S. Treasury tabulations of 4-percent NSLDS sample.
a. Cohorts are defined by the fiscal year when they entered repayment.

IV. Who Are These New Borrowers? Characteristics and Educational Outcomes of Nontraditional Borrowers

The changes in who borrowed and where they borrowed have important implications for the composition and credit quality of the pool of borrowers, their educational outcomes, the amount of debt borrowers accrued, and their economic well-being after enrollment. Compared to other borrowers, the students who borrowed to attend for-profit and 2-year institutions were from more disadvantaged backgrounds (based on their family income) and were older, more independent, and, especially during the recession, more likely to have struggled in the labor market. Nontraditional borrowers tended to attend institutions with relatively poor completion rates, and many appear to have failed to complete the programs they started. These latter factors, in particular, are associated with relatively poor labor market and loan outcomes.

IV.A. Demographics and Family Background

The panels of figure 4 summarize the characteristics and educational outcomes of borrowers entering into repayment in 2011 to provide an

understanding of how changes in enrollment and borrowing patterns affected these borrowers' overall characteristics. In each panel, the figure presents a cross-section of the characteristics or educational outcomes of borrowers who begin to borrow at different types of institutions as undergraduates.[25] While their characteristics have changed within institutions over time, the cross-sectional differences are persistent and therefore give a fairly good indication of how shifts in the share of borrowers across types of institutions are likely to change the characteristics of the borrowing pool.

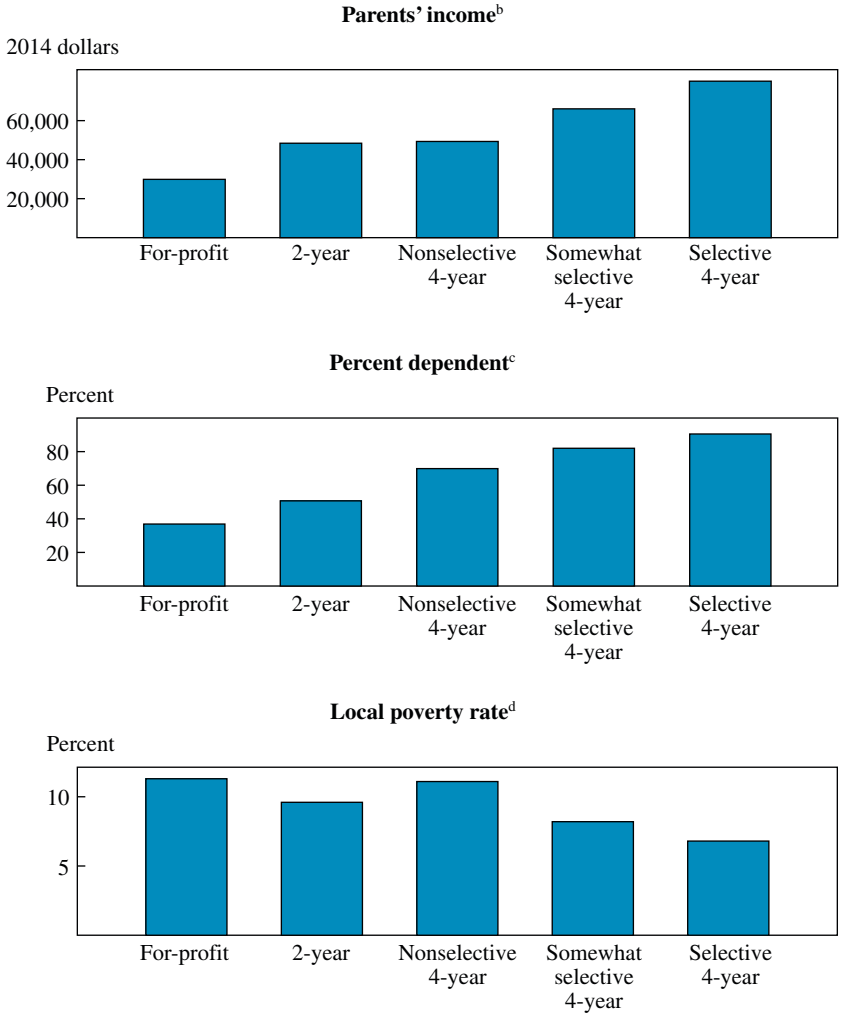
The top-left panel shows that borrowers at more selective institutions tend to come from relatively more affluent backgrounds, with the median family income of dependent borrowers at the most selective institutions being about $80,000, as compared with about $48,000 for borrowers at nonselective 4-year schools and 2-year schools and $30,000 for those at for-profit institutions. These disparities widened modestly over the preceding decade. Because higher family income is positively correlated with labor market outcomes and negatively with default, these differences are one reason why outcomes vary across groups. It is well known that for-profit students tend to come from lower-income backgrounds (Deming, Goldin, and Katz 2012; Cellini 2009), and the observed results are consistent with more vulnerable borrowers from lower-income backgrounds increasingly borrowing to enroll in these institutions.

As the top-right and middle-left panels show, nontraditional borrowers are likely to be older and independent for financial aid purposes. For the 2011 cohort, the median age at entry was 24 for for-profit schools and 23 for 2-year institutions, compared to 19 for students at 4-year institutions; for graduate-only borrowers, the median age at entry was 26. More than 90 percent of borrowers at the most selective 4-year schools were dependent borrowers as compared with 70 percent at nonselective institutions, 50 percent at 2-year institutions, and 37 percent at for-profit schools. This pattern is important for two reasons. First, older, independent borrowers may have less ability to draw on their families for support during times of hardship. Second, independent borrowers have substantially higher loan limits, which allow them to take out more loans each year and to accumulate a higher total amount.

The middle-right and bottom-left panels illustrate two other dimensions of disadvantage. The middle-right panel shows the fraction of borrowers that are reported to be first-generation college students based on the information in their financial aid forms. About 57 percent of the 2011 repayment cohort who had attended for-profit schools were first-generation

25. Graduate-only borrowers are excluded.

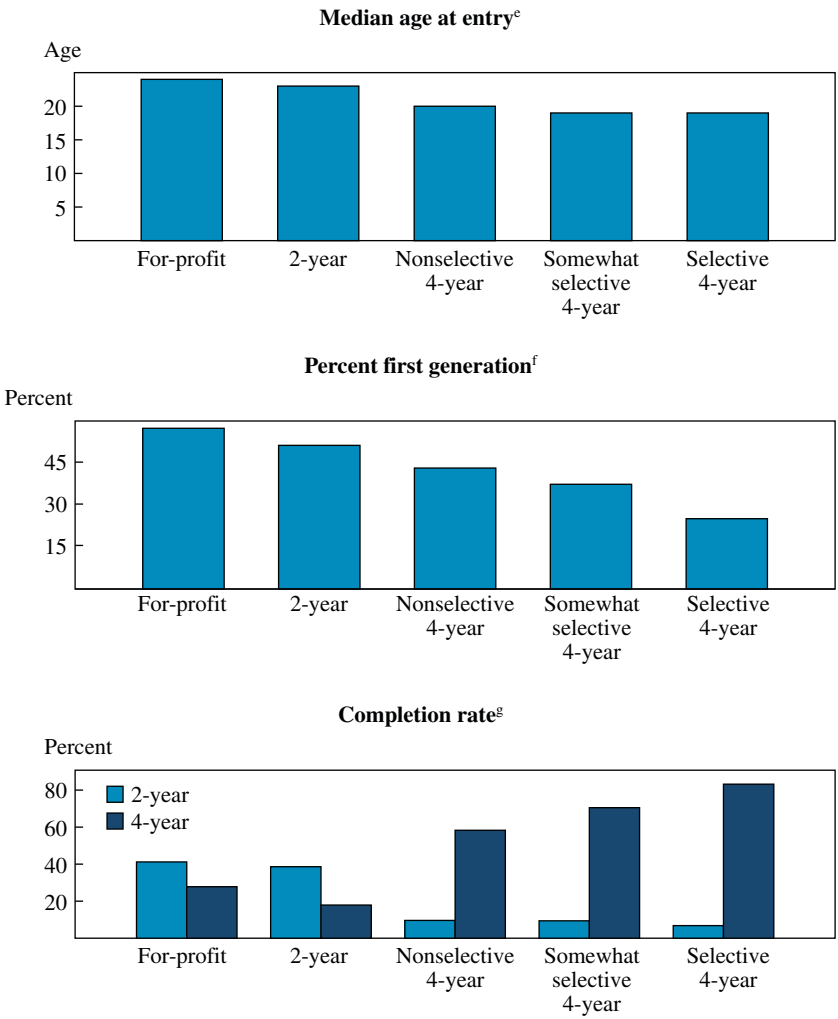
Figure 4. Characteristics of Student Borrowers by Institution Type, 2011[a]



(*continued*)

postsecondary students, compared to 51 percent of those at 2-year schools, 43 percent at nonselective 4-year institutions, and 25 percent at the most selective institutions. Similarly, students from 2-year institutions and for-profit schools were more likely to live in areas with a higher fraction of households living in poverty and with a higher minority population (based on the 2000 Census).

Figure 4. Characteristics of Student Borrowers by Institution Type, 2011[a] (*Continued*)



Source: U.S. Treasury tabulations of 4-percent NSLDS sample. Institution-reported degree completion of borrowers comes from the U.S. Department of Education.

a. This figure examines how changes in enrollment and borrowing patterns at different institutions affect the characteristics of student borrowers in the aggregate. Each panel presents a cross-sectional comparison of borrowers who started borrowing in 2011 as undergraduates at each institution type.

b. Median family (parents') income of dependent undergraduate borrowers.

c. Percentage of borrowers deemed dependent undergraduates in their financial aid applications in the year first borrowed.

d. Average poverty rate in each borrower's ZIP code based on the 2000 Census and the first application for federal aid.

e. Median age of borrowers in the year they first borrowed.

f. Percentage of borrowers reporting that neither of their parents had completed postsecondary education on their initial application for federal aid.

g. Average institution-reported degree completion of borrowers, by degree type.

IV.B. Educational Outcomes

Finally, the bottom-right panel in figure 4 shows that nontraditional borrowers and those from nonselective 4-year institutions appear less likely to graduate from their programs of study than borrowers from most 4-year public and private institutions, based on information reported by the institution to the NSLDS.[26] For instance, at 4-year public institutions, about 71 percent of borrowers completed a 4-year degree, as had 83 percent of borrowers at 4-year private institutions, as reported by the institutions to NSLDS. Among borrowers entering repayment in 2011 who had started at for-profit institutions, about 28 percent had completed a 4-year degree and 41 percent a 2-year degree. (Among borrowers who had started at 4-year for-profit institutions, about 49 percent had completed a 4-year degree and 6 percent a 2-year degree.) At 2-year public institutions, about 39 percent of borrowers were reported to have completed a 2-year degree and 18 percent a 4-year degree.[27] In other words, the institution of first enrollment also had important implications for whether these borrowers were likely to complete a degree and whether they would earn a 2-year or 4-year degree. As we show later, completion rates are strongly associated with subsequent student loan default.

In all, the rise of nontraditional borrowing shifted the composition to borrowers more likely to struggle with their loan burdens—toward older, mid-career borrowers; borrowers from more disadvantaged family backgrounds and poorer neighborhoods; and toward programs many were less likely to complete.

V. Labor Market Outcomes of Borrowers

Other key differences between traditional and nontraditional borrowers are their divergent labor market outcomes and the differential impact the recession has had on each group. Drawing on earnings records from tax data,

26. Because completion and withdrawal measures are solely reported by institutions to the NSLDS for purposes of determining the date at which a loan must enter repayment, institutions have no incentive to ensure the accuracy of their reports. Thus it makes no difference whether an institution reports that a student withdrew or graduated, and some institutions appear to report a student as having withdrawn even if the student graduated. Nevertheless, these completion measures are highly correlated at the institution level with better validated data and, in our analysis, are highly correlated with other outcome measures, like default rates.

27. Some students who begin their postsecondary education at a 2-year school complete 4-year degrees at another institution.

we examine the labor market outcomes of these borrowers after they have entered repayment. For each repayment cohort, the labor market outcomes of borrowers differed based on the institutions they attended, with traditional borrowers earning substantially more. In addition, the earnings and employment rates of nontraditional borrowers declined much more over time, particularly during the recession.

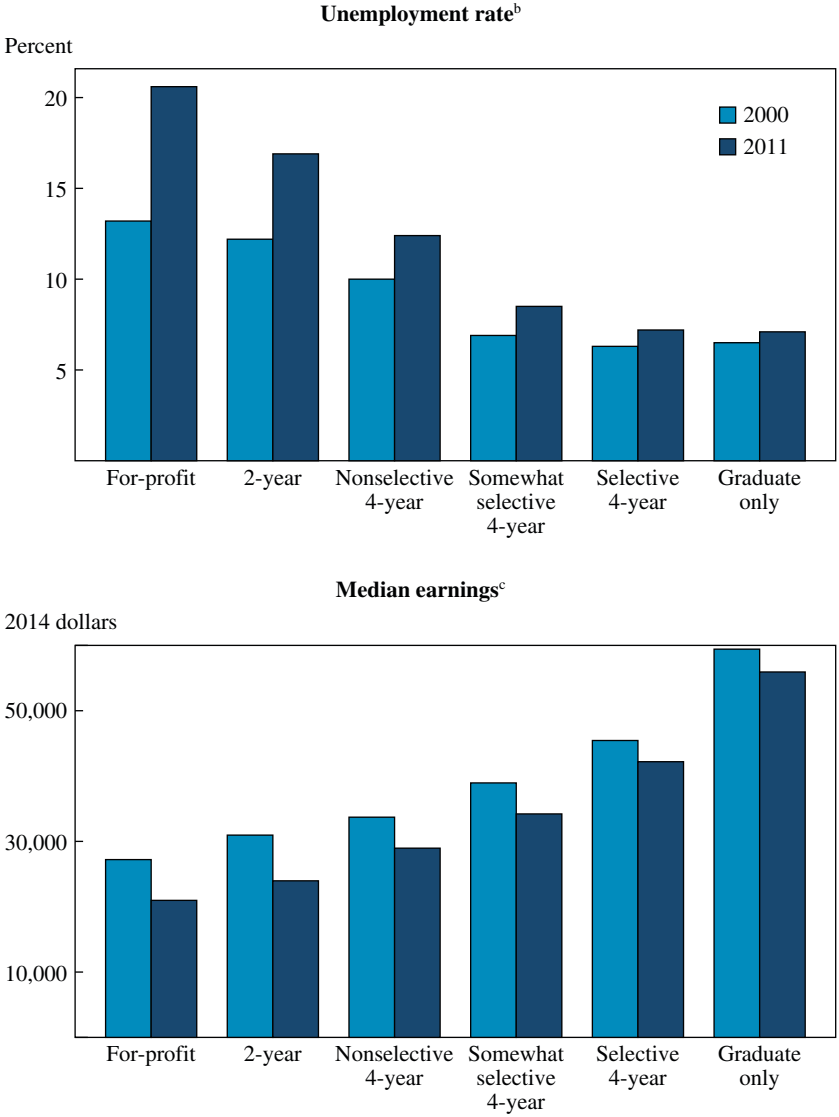
Focusing first on the outcomes of all borrowers in repayment, including those who may have entered repayment many years earlier, shows surprising strength in the outcomes of student loan borrowers. Between 2002 and 2013, the median earnings of all employed traditional borrowers in repayment actually increased. For graduate borrowers, the increase was from $61,000 to $63,100; for borrowers from the most selective 4-year institutions, from $47,300 to $48,000. During the same period, median earnings *declined* among nontraditional borrowers. The declines differed across sectors: from $24,800 to $23,200 for for-profit borrowers and from $30,100 to $25,900 for 2-year borrowers. Unemployment rates edged up slightly (by about 1 percentage point) among all groups.[28]

Focusing more narrowly on borrowers in their earliest years of repayment, we find that while the pattern of relative outcomes is similar, all groups fared worse. The top panel of figure 5 shows the unemployment rate of borrowers by institution type and compares the experiences of the 2000 and 2011 cohorts two years after entering repayment. For the 2000 cohort, unemployment rates among traditional borrowers were low, ranging from 6.3 percent for borrowers from selective 4-year schools to 6.5 percent for graduate borrowers and 10 percent for borrowers from nonselective 4-year schools. For the same cohort, the unemployment rate for 2-year borrowers was 12.2 percent and 13.2 percent for for-profit borrowers. Hence, even prior to the recent recession, there were large differences in employment across borrowers by institution type.

During the recession, unemployment rates rose substantially for nontraditional borrowers, but they rose much less among other borrowers. Among for-profit borrowers the rate jumped to 20.6 percent and among 2-year borrowers to 16.9 percent. For relatively selective 4-year borrowers, the rate increased from 6.3 to 7.2 percent, and for graduate borrowers it

28. We define unemployment as having less than $1,000 in earnings in a year, a definition that differs from the official definition of unemployment in that it does not differentiate voluntary nonemployment (being out of the labor market or not looking for work) from involuntary unemployment. In addition, this measures whether a borrower was employed at all during a calendar year rather than during a particular week.

Figure 5. Unemployment and Median Earnings of Federal Borrowers in Second Year of Repayment, 2000 and 2011 Cohorts[a]



Source: U.S. Treasury tabulations of 4-percent NSLDS sample matched to de-identified tax records.

a. Cohorts are defined by the fiscal year that they entered repayment.

b. Percent of unemployed borrowers (defined here as having less than $1,000 in annual earnings) two years after having entered repayment on all federal loans.

c. Median earnings in 2014 dollars of employed borrowers (defined here as having at least $1,000 in annual earnings) two years after having entered repayment on all federal loans. Earnings defined as W-2 reported wage income plus deferred compensation plus any earnings reported on Schedule SE.

increased from 6.5 to 7.1 percent. In other words, even among students leaving school in 2010 and 2011, near the peak of the recession, there was almost no change in the rate of employment among most traditional borrowers. While the insulating effects of a college degree are apparent in the aggregate unemployment statistics, it is clear that those effects also applied to even most young college borrowers in the years immediately after enrollment.

For those who did find work, a similar pattern applies to their earnings. The bottom panel of figure 5 presents the median earnings of borrowers with earnings of at least $1,000 by the institution type they first attended for the 2000 and 2011 repayment cohorts. In both cohorts, graduate-only borrowers and borrowers from more selective 4-year institutions earned substantially more than other borrowers. For the 2011 cohort, for instance, the median graduate-only borrower earned about $56,100 and the median borrower from a selective undergraduate institution earned about $42,300. In contrast, the median for-profit borrower who worked earned about $20,900 and the median borrower from a 2-year institution about $23,900. The median borrower from nonselective 4-year institutions earned about $29,100.

This pattern reflects two things. First, it reflects long-standing differences in earnings levels across borrowers from different institutions; even in 2000, a borrower from a 4-year selective school earned roughly 66 percent more than a borrower from a for-profit school. Second, it reflects the disproportionate blow to the labor market outcomes of borrowers from less-selective institutions, for-profit schools, and 2-year colleges. Between the 2000 and 2011 cohorts, the median earnings of borrowers declined by 24 percent among for-profit borrowers (two years after entering repayment), 23 percent among 2-year borrowers, and 14 percent among nonselective 4-year borrowers, but only 7 percent among borrowers from the most selective institutions, and 6 percent among graduate borrowers.

Nontraditional borrowers entering the labor market in 2011 therefore faced a particularly severe outlook, with almost 20 percent of them unemployed and with the earnings of those who were working down more than 20 percent relative to their peers in earlier years. Hence, while most federal borrowers in repayment on their outstanding student loans had experienced relatively little change in earnings and employment over the course of the recession (or at least, had roughly the same earnings and employment rates as their peers in earlier years), the most recent cohorts of students faced particularly unfavorable outcomes.

VI. Debt Burdens

The previous sections illustrate that there were many more nontraditional borrowers during and after the recession, that they are a particularly disadvantaged and high-risk group, and that they face relatively poor labor market outcomes when finishing school, particularly the most recent cohorts of borrowers exiting school during and after the recession. This section examines how much they owed in federal loans and their debt service burdens.

VI.A. Increases in Borrowing per Student

Nontraditional borrowers accrued relatively less debt than other borrowers. As table 6 shows, among borrowers entering repayment in 2011, the median borrower at for-profit institutions entered with about $10,500 in debt and the median borrower at 2-year institutions with about $9,600, compared to median borrower debts of $17,600 among those at nonselective 4-year public and private institutions and $23,000 among those at the most selective institutions. Nevertheless, the increases in per-borrower debt have been much larger among nontraditional borrowers and borrowers from nonselective schools.

While the pattern of increases in average balances is largely the same across institutions, the magnitude of the increase is much larger among nontraditional borrowers because a rising share of them accumulated substantial loan burdens. For instance, among borrowers who started at for-profit institutions, the increase in average balances was 51 percent between 2000 and 2011 (from $10,700 to $16,200). This compares during the same period with increases of 32 percent for those at 2-year institutions (from $13,000 to $17,100) and 42 percent for those at the most selective 4-year institutions (from $27,500 to $39,100). (These balances reflect the cumulative balance over a student's career, which may include starting to borrow at a 2-year institution but subsequently borrowing at a 4-year institution, or borrowing initially at a 4-year institution and later in graduate school.) Increases in loan limits appear to have accelerated the accumulation of federal debt burdens in recent years. For instance, there is a clear increase in borrowing starting after the 2007–08 academic year, when loan limits were raised by changes made under the Higher Education Reconciliation Act of 2005.

VI.B. Repayment Burdens Relative to Earnings

The combination of higher loan amounts and worsening labor market outcomes has increased the burden on borrowers. To examine this burden,

Table 6. Median Federal Student Loan Debt in Year Entered Repayment by Institution Type, 1985–2014[a]

Repayment year	*Total*	*For-profit*	*2-year*	*Non-selective 4-year*	*Somewhat selective 4-year*	*Selective 4-year*	*Graduate only*[b]
1985	5,282	4,099	4,148	7,430	7,174	8,113	12,393
1986	4,906	4,056	4,095	7,379	7,228	8,018	12,547
1987	4,757	3,906	3,961	7,127	7,330	8,138	12,663
1988	4,520	3,916	3,971	6,267	7,188	8,490	12,485
1989	5,307	3,846	3,859	6,529	6,965	8,826	13,886
1990	5,861	4,458	4,064	6,211	6,924	9,141	14,545
1991	5,650	4,007	4,049	6,119	6,921	9,460	16,557
1992	6,014	3,921	3,946	6,516	7,374	10,417	16,835
1993	6,734	3,841	4,089	6,826	8,101	11,067	19,936
1994	7,292	4,681	4,400	7,122	8,491	11,750	20,235
1995	8,590	5,750	4,891	8,070	9,911	13,224	22,008
1996	9,865	6,304	5,643	9,124	11,772	15,388	25,291
1997	11,462	6,684	6,206	10,504	14,133	17,856	28,909
1998	12,940	7,316	6,628	12,057	15,394	19,358	30,733
1999	13,865	7,402	7,187	12,196	16,764	20,653	32,970
2000	13,942	7,526	7,125	12,812	16,929	20,575	33,272
2001	14,359	7,756	7,150	13,403	17,961	20,536	34,063
2002	14,369	7,534	7,127	13,210	18,167	20,331	33,797
2003	14,235	7,546	6,881	13,084	18,058	20,584	33,670
2004	13,806	7,346	6,709	13,342	18,113	20,527	32,573
2005	14,534	7,416	7,076	14,094	18,363	20,182	36,579
2006	14,714	7,689	7,277	14,963	18,924	20,494	37,370
2007	13,171	7,594	7,212	14,375	18,359	19,847	34,913
2008	13,504	7,775	7,529	14,712	18,186	19,662	37,874
2009	13,587	8,567	7,956	14,850	18,008	19,128	38,176
2010	14,829	10,162	9,379	17,468	20,364	20,585	41,495
2011	15,265	10,482	9,590	17,588	21,816	22,921	39,460
2012	16,883	11,447	9,861	18,851	23,467	24,767	40,945
2013	18,333	12,693	10,435	20,147	25,013	26,459	42,137
2014	19,647	14,255	11,701	21,229	25,886	26,491	45,890

Source: U.S. Treasury tabulations of 4-percent NSLDS sample. Selectivity data come from Barron's Educational Series (see text).

a. Median total debt burdens in 2014 dollars of students in each fiscal year required to make first loan payments (generally 6 months after leaving school). School types defined by the institution first borrowed to attend. Balances may include both undergraduate and graduate debt.

b. Refers to borrowers who started borrowing at the graduate level.

Table 7. Median Debt-Service-to-Earnings Ratio, Two Years after Entering Repayment, 1999–2010[a]

Repayment year	*Total*	*For-profit*	*2-year*	*Non-selective 4-year*	*Somewhat selective 4-year*	*Selective 4-year*	*Graduate only*[b]
1999	5.3	3.8	3.5	5.3	6.4	6.5	8.4
2000	5.3	3.9	3.3	5.5	6.3	6.6	8.1
2001	5.7	4.2	3.5	6.0	7.0	6.7	8.7
2002	5.4	3.9	3.2	5.5	6.6	6.2	7.7
2003	4.9	3.6	2.9	4.9	6.0	5.8	7.1
2004	4.6	3.4	2.7	4.8	5.7	5.3	6.6
2005	4.8	3.4	2.9	5.0	5.6	5.2	7.5
2006	5.2	3.9	3.3	5.7	6.1	5.6	7.8
2007	5.5	4.5	3.7	6.4	6.8	6.2	8.2
2008	5.9	4.8	4.0	6.9	7.2	6.4	8.9
2009	6.0	5.7	4.4	6.9	7.0	6.1	9.1
2010	6.8	6.8	5.2	8.0	7.9	6.6	10.3
2011	7.1	6.9	5.5	8.3	8.7	7.5	9.9

Source: U.S. Treasury tabulations of 4-percent NSLDS sample matched to de-identified tax records. Selectivity data come from Barron's Educational Series (see text).

a. Debt service estimated using 10-year amortizing loan and (weighted average) interest rate on student balances. Institution types defined as the first institution borrowed to attend. Ratios shown as percentages. See table 6 notes for additional definitions.

b. Refers to borrowers who started borrowing at the graduate level.

we produce estimates of debt-service-to-earnings (DE) ratios. To provide consistent measures of debt service, we assume the standard 10-year repayment plan (a 10-year amortizing loan) and use the (weighted average) interest rate on each student's loans in the year of repayment to estimate the annual payments.[29]

Table 7 provides estimates of the median DE ratio (the median debt service payment divided by median earnings of employed borrowers) for borrowers two years after entering repayment, by institution type. For the cohort entering repayment in 2011, the overall DE ratio was approximately 7.1 percent, almost two percentage points above the ratio of 5.3 percent in 2000. DE ratios have edged up within all groups since 2000, from 3.9 to 6.9 percent among for-profit borrowers, from 3.3 to 5.5 percent among

29. Note that the debt service burdens we calculate here differ from those used for the Department of Education's "Gainful Employment" regulations, which use a 10-year amortizing schedule for less-than-4-year degrees or certificate programs, a 15-year schedule for bachelor's degree programs, and a 20-year schedule for graduate programs. Hence, the debt-service-to-earnings ratios we calculate will be higher and not comparable to those used in the rule.

2-year borrowers, and from 8.1 to 9.9 percent among graduate borrowers. Borrowers from selective institutions experienced the smallest increase of 0.9 percentage point. It is important to note that these ratios compare debt burdens to earnings, but not to ability to pay. Hence, the median annual debt service payment we calculate for the 2011 cohort of for-profit borrowers ($2,200) or 2-year borrowers ($2,300) may be a much larger share of their disposable income than the annual payments for borrowers from relatively more selective 4-year institutions ($5,400) or graduate-only borrowers ($8,700).

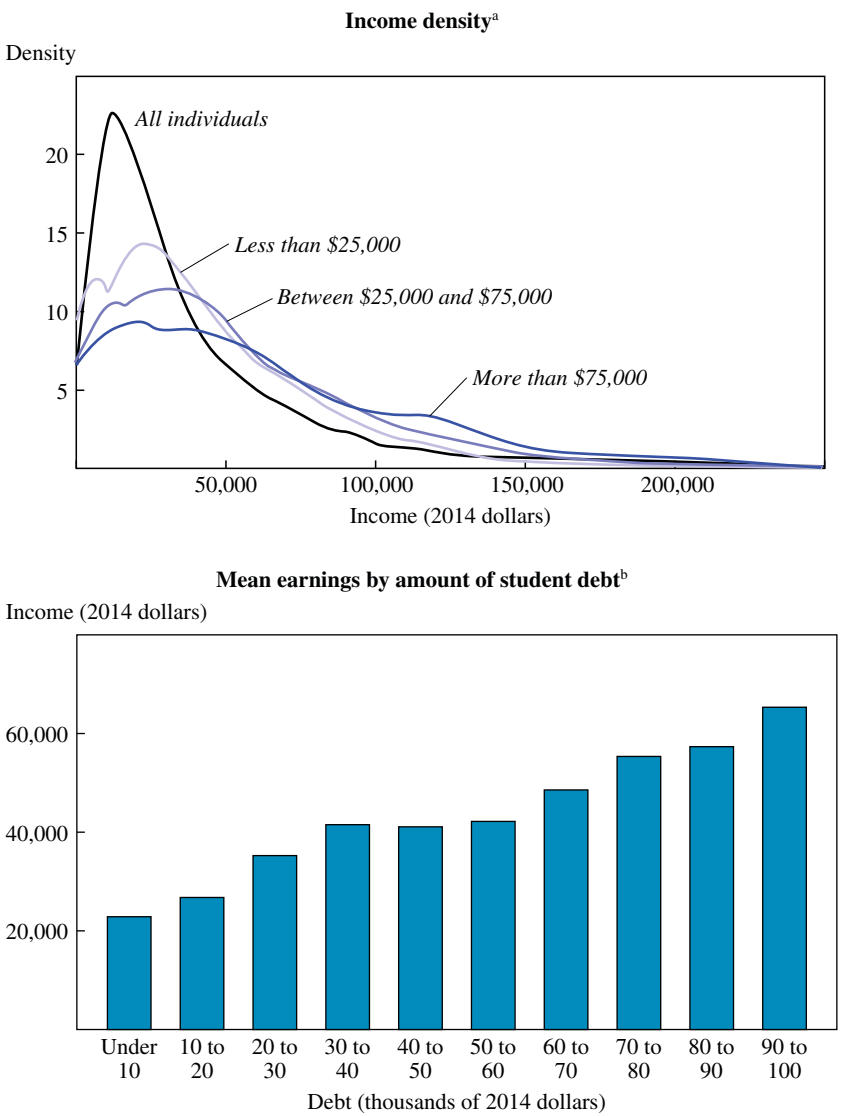
Borrowers from 2-year colleges have historically had the lowest DE ratios—around 3 percent prior to the 2007 cohort—although, as in other sectors, their DE ratios have risen as well. While the median DE ratio for 2-year borrowers increased from 3.3 percent for the 2006 cohort to 5.5 percent for the 2011 cohort, the ratio remained below that for other institution types. Borrowers at for-profit institutions historically had relatively low DE ratios, but saw the largest increases during the prior decade, rising by 3 percentage points from 2000 to 2011. DE ratios tend to be higher early in repayment, which is also when the majority of defaults tend to occur, and fall gradually over time as borrowers' nominal earnings rise during repayment. It is worth noting that the sectors driving the vast majority of defaults—the for-profit and two-year sectors—do not have higher DE ratios than other sectors. In fact, graduate borrowers have the highest DE ratios but the lowest default rates.

VI.C. The Distribution of Debt by Income Level

Figure 6 illustrates the relationship between the amount of debt a borrower has accrued and his or her earnings. The upper panel shows nonparametric estimates of the density of earnings.[30] The lightest line shows the earnings density for borrowers with less than $25,000 in debt, and the moderately dark line shows the earnings density for borrowers with between $25,000 and $75,000 in debt. The darkest line shows the earnings density for borrowers with a large amount of borrowing—more than $75,000, which corresponds roughly to the 95th percentile of outstanding debt. The sample is restricted to individuals between the ages of 25 and 34 and individuals with a positive loan amount. As a basis for

30. The kernel density estimate at a point x is given by $\hat{f}_K = \frac{1}{hn}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$, where h is the bandwidth of an Epanechnikov kernel K.

Figure 6. Relationship between Student Debt and Student Earnings



Source: U.S. Treasury tabulations of 4-percent NSLDS sample matched to de-identified tax records.

a. Figure shows kernel density estimates of the earnings distributions of borrowers with different levels of debt and of the total population of earners (both with and without debt). Sample is restricted to individuals between the ages of 25 and 34.

b. Figure shows mean earnings two years after entering repayment by student debt balance.

comparison, the dashed line shows the density of all tax filers between the ages of 25 and 34.[31]

The lower panel shows average income in $10,000 borrowing bins in 2010. The figure shows that the larger the student debt balance, the more the student tends to earn. This relationship is intuitive—students with larger debts tend to have been enrolled longer, achieved higher levels of educational attainment, pursued higher levels of postsecondary education (such as a bachelor's or graduate degree instead of a certificate), and have attended 4-year institutions where borrowing amounts are greatest, which tend to be the more selective 4-year institutions. For these reasons, borrowers with more debt tend to earn much more.

Figure 6 also shows that borrowers tend to have higher earnings than nonborrowers. Individuals between the ages of 25 and 34 with no student debt earn $37,545 on average, while individuals with student debt earn $43,224. In addition, larger debt amounts are strongly correlated with higher earnings, with mean incomes of $51,555 for borrowers with more than $25,000 in debt and $40,612 for borrowers with less than $25,000 in debt. Borrowers with student debt above $75,000 on average earn slightly more than $60,000, and there are significantly more individuals with debt above $75,000 than those with less debt who earn more than $100,000. The online appendix provides additional information on debt balances and the distribution of debt by borrower incomes as well as their family income, including information regarding the characteristics of borrowers with especially large balances.

VII. Default Rates and Repayment

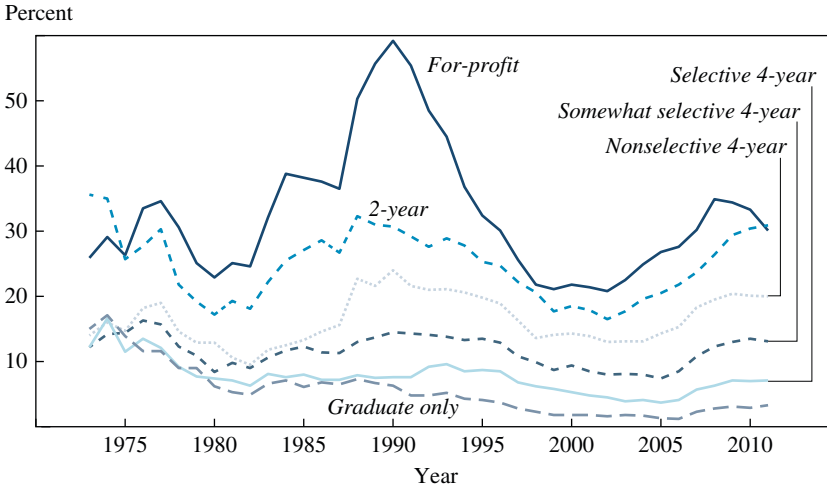
Since its inception in the early 1990s, the 2-year cohort default rate (presented in figure 1) has been the most common indicator of student loan defaults. Cohort default rates are defined by the year in which a cohort enters repayment and are given by the fraction of borrowers who default within a certain number of years after that cohort begins entering repayment. The analysis in this section provides more detail on historical trends in this 2-year measure by institution type and alternative measures of default beyond the 2- and 3-year cohort default rate.

VII.A. Measures of Default and Delinquency

Figure 7 disaggregates the historical 3-year cohort default rate by type of institution first attended. There are large and persistent differences in default

31. The income density is estimated from earnings information and approximated using the log-normal distribution $\mu = 10.1986$, $\sigma = 0.989679$.

Figure 7. Historical 3-Year Cohort Default Rates by Institution Type, 1972–2012[a]



Source: U.S. Treasury tabulations of 4-percent NSLDS sample.

a. The 3-year cohort default rate is defined as the fraction of borrowers entering repayment (on all loans) in each fiscal year who subsequently enter default by the end of the fiscal year two years later. Cohorts are defined by the fiscal year that they enter repayment.

rates across institution type, with nontraditional borrowers experiencing the highest rates of default. Moreover, default rates are most volatile among these borrowers, with default rates rising (and falling) more dramatically among for-profit and 2-year borrowers. Much of the previous 1990 peak in student loan defaults (which led to the introduction of the 2-year cohort default rate rules, the "85/15" rule limiting eligibility to Title IV funds, and automatic wage garnishment of borrowers) was driven by increasing default rates among certain fly-by-night institutions in the for-profit sector (Bennett, Lucchesi, and Vedder 2010), with default rates remaining largely unchanged and even declining in this time period at more selective institutions. Figure 7 shows that over the course of the recent recession, default rates surged among for-profit borrowers, 2-year college borrowers, and borrowers from nonselective 4-year institutions. However, default rates of traditional borrowers increased more modestly and generally remained at or below the levels that prevailed as recently as the mid-1990s.

While default is a salient outcome, it might not capture the increasing loan burdens among borrowers who use alternative payment plans or use forbearance or deferment to suspend payments. Measuring default rates within a certain time frame is important in comparing default rates

between cohorts. However, the relatively short-term cohort default rates exclude outcomes that occur later (or are deliberately deferred outside of a particular default-rate window). The use of forbearance, deferment, and income-based repayment plans is especially relevant following the Great Recession, as take-up has expanded and a large fraction of recent borrowers enroll in these programs. For instance, we estimate that about half of borrowers in recent cohorts have entered into forbearance in their first year of repayment. The use of these programs has helped many students facing labor market challenges avoid default. However, it obscures the fact that many more students may be struggling to repay their loans and that their loan balances remain unpaid, with interest that can continue to accrue.

To examine this, we focus on several alternative measures of loan performance: longer-term default rates, which are defined as the fraction of borrowers who have ever defaulted in a specified period of time after entering repayment; rates of negative amortization, which we define as the fraction of borrowers who owe more in a given subsequent year than they did when entering repayment; and repayment rates, defined as the fraction of total principal and interest a borrower has paid after a given number of years. All of these measures point to deteriorating repayment outcomes, with an especially sharp decline in loan performance among nontraditional borrowers.

The first set of columns in table 8 present the share of borrowers by institution that owe more two years after entering repayment than they did when they first entered repayment. This can occur not only because students defaulted (made no payments but accrued interest) but also in other circumstances where interest is accruing faster than payments are being made. For example, it can occur when a student is in forbearance or deferment and has unsubsidized loans and during the first few years of a graduated repayment plan. It can also occur if, under an income-based repayment plan, the borrower's income is sufficiently low that required payments fall below interest accruals and result in negative amortization. As table 8 shows, rates of negative amortization have surged, with almost three-quarters of for-profit borrowers in the 2012 cohort apparently owing more than when they started borrowing.[32] Rates of negative amortization are also rising within all institution types as more borrowers avail themselves of lower payment options to manage their finances during periods of hardship. These options delay or lower monthly payments, thus alleviating short-term repayment pressure,

32. Prior to 1998, interest accruals on certain loans are missing. Hence, prior to 1996, rates of 2-year negative amortization are lower because they exclude these accruals. The series before and after 1996 are consistently defined.

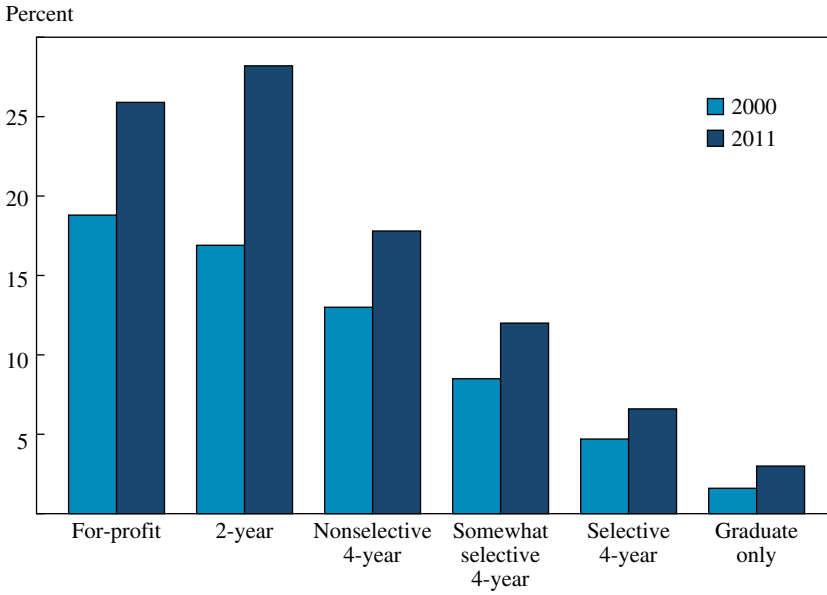
Table 8. Alternative Measures of Loan Performance: Negative Amortization and 5-Year Default Rates, 1986–2012

	Percent of borrowers who owe more two years after entering repayment							*5-year cohort default rates (percent)*						
Repayment year	*All*	*For-profit*	*2-year*	*Non-selective 4-year*	*Somewhat selective 4-year*	*Selective 4-year*	*Graduate*	*All*	*For-profit*	*2-year*	*Non-selective 4-year*	*Somewhat selective 4-year*	*Selective 4-year*	*Graduate*
1986	9	14	12	8	6	4	6	24	41	32	18	14	9	8
1987	9	12	12	8	8	5	6	26	42	32	19	15	10	8
1988	16	22	17	13	8	6	6	37	57	38	27	16	11	9
1989	21	31	19	14	10	6	6	40	62	37	27	17	10	8
1990	23	34	21	17	11	7	7	42	65	37	28	18	10	9
1991	21	32	20	16	10	7	6	38	61	35	26	18	10	7
1992	19	29	19	16	12	9	6	33	56	34	26	18	12	7
1993	16	23	20	16	12	9	8	30	52	36	27	18	12	7
1994	16	22	20	18	13	11	10	25	43	33	26	17	11	6
1995	17	22	22	19	15	12	11	22	38	30	24	17	11	5
1996	33	37	38	37	32	28	29	20	35	29	22	15	10	4

1997	37	42	41	41	35	30	31	18	33	28	20	14	9	4
1998	33	42	39	36	31	26	27	17	30	26	18	13	8	4
1999	34	43	39	38	31	25	26	16	29	24	19	12	8	3
2000	34	43	39	40	31	25	25	16	30	24	19	20	7	3
2001	33	42	37	39	31	25	26	16	30	23	19	12	6	2
2002	33	44	37	37	30	25	23	16	31	23	18	11	6	2
2003	33	44	37	37	29	23	23	17	33	25	18	11	6	2
2004	34	48	40	38	30	23	22	18	36	27	19	11	6	2
2005	36	52	43	40	30	23	21	18	36	27	19	10	5	2
2006	39	54	45	43	33	24	22	19	37	28	20	11	5	2
2007	42	56	46	45	35	26	23	24	40	32	24	15	8	4
2008	44	61	48	45	36	25	25	27	47	35	27	17	9	5
2009	47	66	53	48	38	27	28	28	47	38	27	18	10	5
2010	53	73	59	54	43	31	32							
2011	56	72	62	56	46	34	34							
2012	57	74	64	59	48	36	36							

Source: U.S. Treasury tabulations of 4-percent NSLDS sample. Selectivity data come from Barron's Educational Series (see text).

Figure 8. Increases in 3-Year Cohort Default Rates by Institution Type, 2000 and 2011 Cohorts[a]



Source: U.S. Treasury tabulations of 4-percent NSLDS sample.

a. Cohorts are defined by the fiscal year that they entered repayment. See notes to figure 7 for definitions of default rate.

though interest continues to accrue, and potential financial liabilities of both the student and the federal loan program continue to grow.

The right-hand side of table 8 presents a longer-term perspective on default rates by showing 5-year cohort default rates by institution type. For the 2009 cohort, 47 and 38 percent of borrowers at for-profit and 2-year public and private institutions, respectively, defaulted within 5 years. This is a sharp increase from the 1999 cohort, in which 29 and 24 percent of for-profit and 2-year borrowers defaulted. The 5-year cohort default rates increased between the 2009 and 1999 cohorts at all institution types, but the increase was much smaller at more selective institutions. For example, at the most selective institutions, 5-year cohort default rates increased from nearly 8 percent for the 1999 cohort to 10 percent for the 2009 cohort.

Figure 8 compares the 3-year default rate for the 2000 repayment cohort to that of the 2011 cohort. It illustrates that for-profit, 2-year, and nonselective schools have higher default rates than other institutions. While default rates have increased at all types of institutions, the increases have

been greatest at for-profits and 2-year institutions. This figure is the starting point for the analytical exercise that follows, which attempts to quantify the contribution of the changes described above.

VII.B. Why Have Default Rates Increased?

FACTORS ASSOCIATED WITH DEFAULT How have changes in the pool of borrowers, the institutions they attended, and their educational and labor market outcomes contributed to rising default rates? To help answer this question, we draw on Oaxaca-Blinder decomposition methods to quantify how changes in the observable characteristics of borrowers affected default rates. In this approach, we first estimate basic models of default that represent the 3-year cohort default rate as a function of the type of institution first attended, the characteristics of borrowers, and their post-schooling outcomes. We then use those models to determine how the overall default rate would change holding fixed the model parameters—that is, the likelihood of default conditional on the characteristics of borrowers—but applying the model to borrowers in an alternative year.

For example, consider a simple decomposition of the 9.7-percentage-point increase in the 3-year cohort default rate between 2000 and 2011. Over that period, the share of borrowers who were nontraditional increased from roughly 30 to 48 percent, and nontraditional borrowers in 2011 were approximately 19 percentage points more likely to default than other borrowers. Holding fixed that 19-percentage-point difference, this implies that the change in the share of nontraditional borrowers increased overall default rates by about 3.4 percentage points, or 35 percent of the total 9.7-percentage-point increase. The remaining increase is attributable to increases in default rates within groups—the 10-percentage-point increase in default rates among nontraditional borrowers and the 4-percentage-point increase among traditional borrowers.

Of course, this simple example ignores any concurrent shift in the institutions borrowers attend, the characteristics of borrowers, their educational attainment, or their labor market outcomes. In the remainder of this section, we outline a decomposition approach to examine a variety of observable factors associated with student loan default using a logit model of the following form:

$$D_i^* = \alpha_i + \beta X_i + \gamma Z_i + \varepsilon_i, \tag{1}$$

where D_i is an indicator of default and D_i^* is an unobserved latent variable such that if $D_i^* > 0$, $D_i = 1$, otherwise $D_i = 0$. The term α_i is an indicator for

institution type and school; X_i captures characteristics determined before enrollment, such as family income or dependency status; and Z_i captures characteristics determined after enrollment, such as labor market earnings or duration. The term ε_i is a standard logistic distribution error term, which is assumed to be orthogonal to the outcome conditional on the observables Z_i and X_i. The i subscript denotes the individual student. We assume that $\Pr(D_i = 1 \mid X) = \Lambda(X'\beta)$, where Λ is the cumulative distribution function of the standard logistic distribution.

The model and results are presented in table 9. The characteristics X_i, determined before enrollment, are listed in table 9, and include dependency status, age, gender, family income, marital status, and whether the borrower was a dependent or a Pell Grant recipient. Characteristics Z_i, determined after enrollment, are also listed in table 9, and include duration of enrollment at each institution type, whether the borrower was ever a graduate student, and log income and earnings. The sample is limited to graduate and undergraduate borrowers who entered repayment within 10 years of starting school (to exclude unusual enrollment patterns and parent borrowers). Further discussion of variable construction and data sources is provided in online appendix A.

The first three columns of table 9 pool the entire sample, while columns 4 through 7 split the sample into the 2000 and 2011 cohorts. There is a strong relationship between the type of school attended and loan default, with nontraditional borrowers substantially more likely to default than other borrowers, as indicated by the larger coefficients on for-profit and 2-year indicators. Among traditional borrowers, the selectivity of institutions is strongly associated with default, with borrowers from nonselective schools experiencing default probabilities closer to those of nontraditional borrowers than to those of borrowers from selective institutions.

Even when controlling for observables, the relationship between school type and default remains substantial (and statistically significant at the 1 percent level), but the coefficients drop significantly when individual controls are added, as seen in columns 2 and 3. This is consistent with the evidence above that the rising enrollment of higher-risk borrowers is concentrated in the for-profit and 2-year sectors, at least part of the relationship between default and school type being driven by selection. For instance, students at for-profits and 2-year colleges tend to come from lower-income backgrounds, tend to earn less after graduation, and have weaker academic backgrounds compared to other students. In other words, because attending a for-profit or 2-year institution is correlated with other characteristics that increase default risk, the coefficient on the for-profit indicator pools both

Table 9. Factors Associated with Default[a]

	(1) *Pooled*	*(2)* *Pooled*	*(3)* *Pooled*	*(4)* *2000*	*(5)* *2011*	*(6)* *2000*	*(7)* *2011*
For-profit	0.460***	0.340***	0.131***	0.331***	0.346***	0.130***	0.130***
	(0.0132)	(0.0139)	(0.0136)	(0.0266)	(0.0163)	(0.0234)	(0.0172)
2-year	0.532***	0.446***	0.185***	0.338***	0.479***	0.140***	0.206***
	(0.0140)	(0.0156)	(0.0172)	(0.0286)	(0.0181)	(0.0263)	(0.0219)
Nonselective 4-year	0.431***	0.373***	0.174***	0.282***	0.404***	0.144***	0.186***
	(0.0152)	(0.0159)	(0.0170)	(0.0256)	(0.0193)	(0.0247)	(0.0221)
Somewhat selective 4-year	0.296***	0.286***	0.143***	0.205***	0.316***	0.110***	0.156***
	(0.0144)	(0.0146)	(0.0151)	(0.0204)	(0.0188)	(0.0197)	(0.0204)
Selective 4-year	0.164***	0.167***	0.076***	0.123***	0.181***	0.067***	0.075***
	(0.0137)	(0.0137)	(0.0140)	(0.0175)	(0.0184)	(0.0174)	(0.0194)
Age[b]		−0.003***	−0.002***	−0.002***	−0.004***	−0.001***	−0.002***
		(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Female		−0.063***	−0.051***	−0.046***	−0.071***	−0.038***	−0.057***
		(0.0019)	(0.0018)	(0.0025)	(0.0026)	(0.0023)	(0.0025)
Married[c]		−0.091***	−0.069***	−0.072***	−0.096***	−0.053***	−0.075***
		(0.0032)	(0.0032)	(0.0039)	(0.0046)	(0.0038)	(0.0045)
Has children[d]		0.086***	0.064***	0.072***	0.088***	0.050***	0.066***
		(0.0057)	(0.0051)	(0.0087)	(0.0074)	(0.0074)	(0.0068)
Family income under $5,000[e]		0.127***	0.071***	0.075***	0.153***	0.049***	0.084***
		(0.0037)	(0.0033)	(0.0052)	(0.0048)	(0.0046)	(0.0044)
Family income $5,000–$25,000[e]		0.062***	0.045***	0.032***	0.078***	0.025***	0.056***
		(0.0025)	(0.0024)	(0.0033)	(0.0034)	(0.0030)	(0.0032)
Dependent		−0.028***	−0.021***	−0.014***	−0.034***	−0.010***	−0.026***
		(0.0025)	(0.0024)	(0.0032)	(0.0034)	(0.0030)	(0.0033)
Pell Grant recipient		0.030***	0.007***	0.017***	0.035***	0.000	0.011***
		(0.0022)	(0.0020)	(0.0033)	(0.0028)	(0.0027)	(0.0027)
Has graduate loans			−0.054***			−0.041***	−0.059***
			(0.0070)			(0.0073)	(0.0106)

(*continued*)

Table 9. Factors Associated with Default[a] (*Continued*)

	(1) *Pooled*	*(2)* *Pooled*	*(3)* *Pooled*	*(4)* *2000*	*(5)* *2011*	*(6)* *2000*	*(7)* *2011*
Earnings[f]			–0.051***			–0.029***	–0.062***
			(0.0014)			(0.0017)	(0.0019)
Income[f]			–0.017***			–0.011***	–0.020***
			(0.0008)			(0.0011)	(0.0011)
Years of attendance[g] *at*							
For-profit			–0.014***			–0.004	–0.018***
			(0.0017)			(0.0029)	(0.0022)
2-year			0.003			0.000	0.003
			(0.0019)			(0.0029)	(0.0025)
Nonselective 4-year			–0.003			–0.005*	–0.002
			(0.0016)			(0.0020)	(0.0023)
Somewhat selective 4-year			–0.006***			–0.004*	–0.007***
			(0.0015)			(0.0018)	(0.0021)
Selective 4-year			–0.008***			–0.005*	–0.010***
			(0.0019)			(0.0021)	(0.0028)
Graduate			–0.005			0.003	–0.012*
			(0.0039)			(0.0042)	(0.0057)
No. of observations	147,770	147,733	147,703	46,774	100,927	46,753	100,918

Source: U.S. Treasury tabulations of 4-percent NSLDS sample matched to de-identified tax records. Selectivity data come from Barron's Educational Series (see text).

a. This table gives results from the logit model presented in-text. The dependent variable is an indicator of whether or not a loan defaults within three years of repayment (a loan is in default if a payment is more than 270 days overdue). The sample is restricted to individuals who entered repayment in 2000 or 2011 who had first started borrowing within the preceding 10 years at the undergraduate or graduate level (excluding parent loans). All specifications include indicator variables for institutions with more than 50 observations in 2011 (about 1,250 aid recipients per cohort), approximately the largest 600 institutions. The omitted school type category is graduate-only borrowers. Additional sample restrictions are given in the header above each column. Statistical significance indicated at the *10 percent, **5 percent, and ***1 percent levels. Huber-White robust standard errors in parentheses.

b. Refers to the age when an individual entered repayment.

c. Refers to filing status two years after entering repayment.

d. Refers to number of dependent children two years after entering repayment.

e. Family income as recorded on the borrower's first FAFSA.

f. Earnings and income are the inverse hyperbolic sine of earnings and total family income (adjusted gross income before adjustments) two years after entering repayment.

g. Years of attendance variables are years of active borrowing at each institution type.

the causal effect of attending a for-profit and unobserved characteristics correlated with both default and attending a for-profit.

Low family income is associated with default (and statistically significant at the 1 percent level). Students from lower-income backgrounds are substantially more likely to default than their peers from higher-income households. Attainment, measured by years of schooling at each institution type, shows a strong relationship with default, as students with more years of schooling generally have a lower risk of default. Accumulating more years of schooling is generally associated with better labor market outcomes and, in addition, the years-of-schooling variables are also a proxy for whether a borrower has dropped out (for example by reflecting a borrower's having attended for only a few years).[33] Labor market outcomes, like earnings and income, are also closely related to default, with higher-income borrowers substantially less likely to default.

Comparing the specifications in columns 4 through 7, the coefficients on each covariate are of the same sign and generally of a similar magnitude, indicating that the relationship between observables and default is similar across years. The basic patterns observed are consistent with the default regressions found by Laura Greene Knapp and Terry Seaks (1992). However, for the 2011 cohort, the effects of most covariates appear larger than in 2000, including family income, educational attainment, labor market outcomes, and school types; this suggests that along these dimensions, borrowers' outcomes diverged further among the 2011 cohort. The increasing effect of school types is consistent with changes in the selection of borrowers and the increasing enrollment of borrowers who are more likely to default on student loans.

DECOMPOSITION ANALYSIS We use decomposition methods to examine how much of the increase in default rates can be explained by changes in the characteristics of students, the institutions they attended, and their labor market outcomes, under the strong assumption that the relationship between these observable characteristics and the likelihood of default remained the same in 2000 and 2011. Under this assumption, we use the models estimated above to predict how default rates would have changed because of changes in these observable characteristics, holding fixed the other parameters of the model, and interpret the predicted increases in default as the amount that can be explained by changes in the characteristics. We use a

33. While institution-reported completion measures are included in the data, completion appears to be underreported in earlier years, making reliable comparisons between 2000 and 2011 difficult.

nonlinear variation on the standard Oaxaca-Blinder decomposition, which offers two significant advantages.[34] First, student loan defaults lie in the tail of the distribution, where linear estimators tend to perform poorly. Second, there are large gaps in explanatory variables between 2000 and 2011, and linear estimators can lead to predicted probabilities above 1 or below 0.

The results of the decomposition do not necessarily have a causal interpretation because the counterfactual depends on whether regression results reflect a causal relationship between particular variables and default. For example, changes in earnings may or may not have had a causal impact on default rates, depending on whether earnings have had a causal impact on default through liquidity constraints or another channel, and depending on whether earnings are correlated with default through unobservable channels such as primary schooling or access to family resources. The strong association between attending a for-profit and defaulting could reflect both the causal impact of attending a for-profit and the effect of unobservably riskier students sorting into for-profits. To date, there has been more descriptive evidence and less analysis of the causal impact of various factors on default, and many questions remain open. This point is important where policy recommendations are concerned. Policies aimed at lowering defaults by affecting observed correlates of default may or may not have their intended effect depending on the causal nature of the relationship between the particular variable and default.

Table 10 presents the results of the nonlinear decomposition.[35] The first row of table 10 shows 3-year cohort default rates in 2000, the second row shows default rates in 2011, and the third row shows the difference. The fourth row shows the change associated with observed explanatory variables (endowments) given by the procedure outlined above. The first and second columns include only indicators for the type of institution attended and school dummies for larger institutions. The first column uses the 2011 coefficients to predict default rates in 2000, while the second column uses the 2000 coefficients to predict default rates in 2011.

The results in columns 1 and 2 indicate that between one-quarter and one-half of the total increase in default is associated with changes in the types of institutions students attend. Columns 3 and 4 add family background and student characteristics (age, gender, marital status, family

34. For further detail on the nonlinear procedure developed by Blinder (1973) and Oaxaca (1973), a detailed description is given in the online appendix.

35. The results are similar using a linear decomposition.

Table 10. Nonlinear Decomposition of Increase in Student Loan Default[a]

	(1) Full sample	(2) Full sample (reversed)	(3) Full sample	(4) Full sample (reversed)	(5) Full sample	(6) Full sample (reversed)	(7) Nontraditional borrowers	(8) Traditional borrowers
2000 cohort default rate	0.125	0.125	0.125	0.125	0.125	0.125	0.219	0.086
2011 cohort default rate	0.222	0.222	0.222	0.222	0.222	0.222	0.318	0.125
Difference[b]	–0.097	0.097	–0.097	0.097	–0.097	0.097	–0.096	–0.039
Endowments[c]	–0.048	0.026	–0.050	0.038	–0.061	0.051	–0.042	–0.013
School type[d]	–0.048	0.026	–0.038	0.024	–0.020	0.011	–0.025	–0.005
Background[e]			–0.012	0.015	–0.006	0.009	0.000	0.000
Education[f]					–0.007	0.009	0.022	0.002
Labor market[g]					–0.028	0.022	–0.039	–0.010
Percent of total difference								
Total	50	26	52	40	63	53	43	33
School type[d]	50	26	39	25	20	11	26	14
Background[e]			13	15	7	9	0	–1
Education[f]					7	10	–23	–5
Labor market[g]					29	23	41	26

Source: U.S. Treasury tabulations of 4-percent NSLDS sample matched to de-identified tax records.

a. See text and table 9 for description of the model covariates and sample selection. In columns 1, 3, 5, 7, and 8, the base year is 2011 and the counterfactual is 2000 using covariates from the 2011 model. In columns 2, 4, and 6, the counterfactual is reversed. Statistical significance indicated at the *10 percent, **5 percent, and ***1 percent levels.

b. Shows the difference between the 2000 and 2011 cohort default rate. For columns 1, 3, 5, 7, and 8, this is the 2000 cohort default rate minus the 2011 cohort default rate; for columns 2, 4, and 6, this is the 2011 cohort default rate minus the 2000 cohort default rate.

c. Shows the difference between simulated default rates using the counterfactual distribution of observables.

d. Includes indicators for type and selectivity of institution attended and institution fixed effects for institutions with more than 50 observations (large schools).

e. Includes characteristics of the borrower, including age, gender, marital status, family income, and whether borrower received a Pell Grant.

f. Includes whether a graduate student and years of attendance at each institution type.

g. Includes inverse hyperbolic sine of earnings and income.

income, and whether the student was a dependent or Pell Grant recipient). The results in these columns indicate that approximately 15 percent of the increase in default is associated with changes in family background, after controlling for institution. Columns 5 and 6 add controls for labor market outcomes (earnings and income), educational outcomes (like duration of enrollment by institution type and whether the borrower has graduate loans), and background characteristics. The results in these columns indicate that approximately one-quarter to one-third of the increase is associated with deteriorating labor market outcomes (conditional on the other factors).

Looking across columns, the association between school types and default drops when individual characteristics are included, which suggests that the school type indicators are capturing unobserved student-specific factors. Changes in family background characteristics over the time period lead to little change in default rates in the aggregate, largely reflecting the fact that family income did not change much on average as borrowing increased among both lower- and higher-income families. While nontraditional borrowers were from poorer families, traditional borrowers tended to be from somewhat higher-income families. The largest contributors are changes in labor market characteristics, which explain roughly one-quarter of the increase in student loan default. Because family background and labor market outcomes are highly correlated with the institutions borrowers attend, it is difficult to distinguish in the aggregate analysis whether changes in default are arising because of changes in where borrowers attend or changes in the characteristics of the borrowers themselves. The decomposition results indicate that, taken together, changes in observable characteristics of borrowers, institutions, and their labor market outcomes are associated with between one-half and two-thirds of the increase in default between 2000 and 2011.

The final two columns of table 10 present the full decomposition separately for nontraditional and traditional borrowers to examine how changes in borrowers' characteristics and outcomes are associated with sector-specific default rates. Changes in labor market outcomes and family background characteristics explain between one-half and two-thirds of the increase in default among nontraditional borrowers. Roughly one-third of defaults appear to arise for reasons unassociated with characteristics of students or outcomes that we observe; these other reasons may include characteristics or quality of their education or institution, as well as unobserved indicators of students' financial hardship, their expectations about future changes to loan programs, and their willingness to pay. At

nontraditional schools, decreases in earnings and increases in borrowing are associated with a substantial portion of the increase in default; however, much of this is offset by changes in education duration and completion variables, which are associated with a decrease in default rates. Among traditional borrowers at 4-year public and private institutions, changes in labor market outcomes and family background explain a smaller portion of the increase in default, changes that are somewhat offset by changes in education and background-related variables that were associated with decreases in default rates.

The final two columns also illustrate that a large portion of the increase in default rates—almost 5.5 percentage points for nontraditional borrowers and 2.6 percentage points for traditional borrowers—cannot be explained using information we observe. In other words, while a large share of the increase in default rates can be associated with changes in institutions, students, and the labor market, the overall default rate is also rising for other unobserved factors, particularly among nontraditional borrowers.

While much of the increase in default is associated with observed changes in the characteristics of borrowers, institutions, and labor market outcomes, a sizable portion is not. Several unobserved characteristics are likely to be important, leading to an underestimate of the role of institutional factors or borrower backgrounds. For instance, David Deming, Claudia Goldin, and Lawrence Katz (2012) find that students' satisfaction with their institutions is an important contributor to default and that satisfaction is lower among for-profit institutions. Similarly, it is possible that other differences in educational quality, school-specific differences in loan counseling, and other characteristics of students, like their financial situation and their employment prospects, may also have changed, contributing to higher rates of default among borrowers.

A limitation of this approach is that it does not identify the cause of any particular explanatory variable's effect on default. For instance, labor market outcomes among nontraditional borrowers could deteriorate because of (unobserved) changes in the characteristics of the borrowers themselves, such as being drawn from lower-skilled groups, or changes in the characteristics of the institutions they attended. For example, the observed effect of earnings on defaults pools several factors: not only the causal effect of earnings on default but also the fact that borrowers with low earnings might have attended programs that they were unlikely to complete or that did not lead to better jobs, as well as economic shocks that disproportionately affected those borrowers. Moreover, it is possible that part of the change in earnings is due to direct effects of the type of school attended, and it is also

possible that the observed effect of schools is partly driven by certain types of students selecting into certain types of institutions.

In summary, a substantial portion of the overall increase in defaults is associated with nontraditional borrowers, both because they are a rising share of all borrowers and because their default rates have increased. Additionally, lower earnings and deteriorating labor market outcomes are associated with a significant fraction of the increase in default rates between 2000 and 2011.

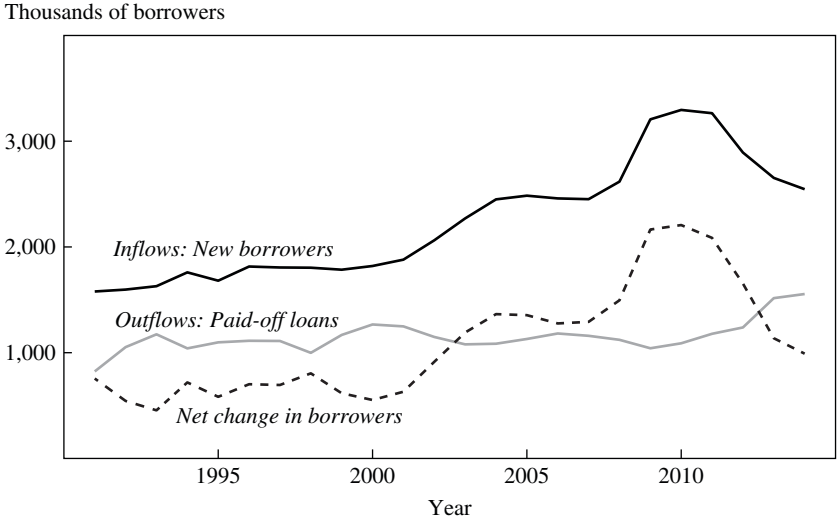
VII.C. The Outlook: Changes in Borrowing Patterns Post-Recession

Borrowing tends to rise in recessions because enrollment increases and borrowing per enrollment increases as well. Educational enrollment is countercyclical, both because adverse economic conditions decrease the opportunity cost of college attendance and because they increase financial pressure due to lower earnings and assets.[36] The Great Recession was no exception to this pattern; during the period, inflows into borrowing increased sharply, and then they declined following the recovery. Moreover, three additional factors related to and coinciding with the Great Recession could have increased borrowing. First, pressure on state budgets led to cuts for many public institutions. Part of these cuts may have been passed on to students, leading to higher borrowing. Second, access was restricted in many alternative credit markets. The restricted availability of other forms of lending may have led to increased student loan borrowing, especially from students who otherwise would have relied on private parental borrowing. Finally, federal initiatives providing information to Unemployment Insurance recipients may have induced many individuals to utilize federal aid and borrowing programs to enroll in college (Barr and Turner 2015a, 2015b). Hence, whether these pressures and patterns are likely to persist has important implications for the long-term costs and benefits of federal lending programs.

Figure 9 shows that the inflow of new borrowers and the outflow of borrowers who paid off loans changed substantially during the recession, but more recently they have begun approaching prerecession levels. Prior to 2002, each year about 1.8 million new borrowers took out loans for the first time and almost 1.2 million borrowers repaid their loans in full; hence the number of outstanding student loan borrowers was increasing by just under 700,000 borrowers a year. Starting in 2003 and continuing through 2007,

36. For example, see Bound, Lovenheim, and Turner (2010).

Figure 9. Changes in the Stock of Borrowers: Flows Into and Out of Indebtedness, 1991–2014



Source: U.S. Treasury tabulations of 4-percent NSLDS sample.

about 2.4 million new borrowers took out first-time loans each year, and over 1.1 million paid off their loans entirely. Hence, the annual increase in the number of borrowers almost doubled to 1.3 million. At the onset of the recession, from 2009 to 2011, the number of new borrowers each year increased sharply, rising to 3.3 million, and the number of borrowers paying loans off entirely edged down to 1.1 million, leading to an increase in the number of borrowers each year of more than 2.1 million.

More recently, the number of borrowers paying off loans has also increased, rising from the recent low of about 1 million in 2009 to 1.6 million in 2014. Led by reductions at both for-profit schools, where first-time borrowing started to drop in 2011, and at 2-year colleges, where it started to drop in 2012, the number of new borrowers fell to 2.5 million in 2014. Given the lag between enrollment, borrowing, and loan repayment, many of the borrowers who enrolled during the Great Recession entered into repayment in 2010 and later (see figure 2).

Just as the sharp rise in new nontraditional borrowers during the Great Recession contributed to rising rates of default, the recent unwinding of recession-related enrollment trends (particularly at for-profit and 2-year institutions), improving economic conditions, and increasing enrollment

in income-based repayment plans are likely to put downward pressure on default rates. When the recession waned, millions of borrowers left school and became responsible for making loan payments for the first time. Hence, not only were there vastly more federal borrowers, but many of them were in their first years of loan repayment, a time when borrowers' careers are just starting and their earnings are the most variable. The wave of borrowers who had begun to borrow during the recession began to start repayment on those loans in increasingly large numbers from 2011 through 2014, when almost 4 million borrowers entered repayment—more than double the number in a typical year prior to the recession. Because borrowers are most likely to default in the first three years of repayment, that wave of borrowers has or will translate into a surge in the number of defaults in 2013 through at least 2015.

However, the life cycle of borrowing makes default a lagging indicator, and the current high rate of delinquency obscures several more favorable recent trends. In particular, the number of new borrowers at for-profit and 2-year institutions has dropped substantially, due to the end of the recession and due to increased oversight of the for-profit sector, which is likely to improve the risk characteristics of future repayment cohorts. That change in composition, together with the slowdown in the number of new borrowers and efforts by the Department of Education to expand and encourage the use of income-based repayment programs, is likely to put downward pressure on loan delinquency in the near future.

Another important metric for gauging the persistence of adverse repayment outcomes is how today's borrowers manage their loan burdens and how much of their burdens are repaid over time. Recent cohorts have been paying off their loan balances at a slower rate than earlier cohorts, with the median borrower in the 2011 cohort actually owing more after two years than he or she owed in the first year of repayment. The typical borrower in cohorts that entered repayment in the late 1990s had repaid her balance within 10 years of entering repayment, but subsequent cohorts have repaid more slowly. This is especially true of nontraditional borrowers. The slowing repayment rates appear to be due to high rates of default in recent cohorts, as well as increased use of programs such as income-based repayment, forbearance, and deferment among unemployed or low-income borrowers, in which borrowers' payments are suspended or reduced. In 2010, 62 percent of borrowers in default appeared to be eligible for income-based repayment on the basis of a means test, which could suggest that many borrowers are unaware of alternative repayment options.

The first observation above (the shift in the composition of borrowers during the recession and the reversal of flows associated with the recession) suggests that more recent borrowers will have better outcomes than those during the recession, if only because the types of borrowers and institutions attended have shifted. But the latter observation (slowing repayment rates) suggests that if recent patterns persist, the burdens owed by yesterday's students may endure for years.

VIII. Concluding Remarks

This paper uses new administrative data sources to examine the characteristics of student loan borrowers experiencing high default rates and the reasons why default rates increased in recent years. We show that high rates of default and other measures of loan delinquency, like the fraction of borrowers failing to make progress repaying loans, are concentrated among nontraditional borrowers. Moreover, since at least 2000, the number of nontraditional borrowers increased rapidly, not only in absolute terms but as a share of federal borrowers and of outstanding balances, particularly in comparison to the number and composition of postsecondary students. More than 30 percent of recent nontraditional borrowers defaulted on their loans within 3 years, and many more are not making progress repaying their loans.

Regression analysis suggests that nontraditional borrowers experience higher rates of default in part because they are drawn from more disadvantaged backgrounds. For instance, nontraditional borrowers were older, more likely to be independent of their parents, from lower-income families, and living in more disadvantaged areas. They borrowed substantial amounts to attend institutions with low completion rates and, after enrollment, experienced poor labor market outcomes that made their debt burdens difficult to sustain.

However, even controlling for borrower characteristics, the institution a borrower attended is strongly associated with his or her loan outcome, suggesting that the relative disadvantage of nontraditional borrowers is insufficient to explain their worsening loan outcomes. How and why educational institutions matter is less clear in our analysis, although important factors are likely to include the quality of the program offered, persistence and completion rates, the program's labor market return, and students' satisfaction. Whatever the fundamental cause, many nontraditional borrowers clearly were not well equipped to succeed in the difficult labor market of the last few years, nor were they able to manage their debt burdens.

Increases in the number of nontraditional borrowers, changes in the institutions they attended, and their worsening labor market outcomes were therefore important contributors to the rise in overall default rates.

In contrast, default rates of traditional borrowers have remained low, and the labor market outcomes of many traditional borrowers have remained stable or even improved in recent years, despite the recession. Traditional borrowers tend to have higher incomes than the general population and to owe larger loan balances. Even traditional borrowers with large balances tend to do well, on average, mainly because they acquired their loan balances while attending selective schools or graduate and professional programs.

While outcomes are likely to improve for today's borrowers, concerns about the student loan program are likely to persist. One concern is that many institutions whose students experience high default rates, low repayment rates, and weak labor market outcomes continue to enroll high-risk borrowers, saddling these students with loans they struggle to pay and leaving taxpayers on the hook for their losses. In the past, policymakers have used institutional accountability measures, like the Cohort Default Rate and "90/10" rules, to reduce student and taxpayer exposure to certain institutions. As the experience in the early 1990s demonstrated, strengthened accountability can reduce defaults. However, such policies have trade-offs, because they may limit the educational opportunities of higher-risk or underserved students. Gauging whether such students (and taxpayers) would be better or worse off from accountability changes or whether policy changes would encourage new and better educational outcomes requires better measures of the returns to educational investments at different institutions.

A related concern arises from recent work that finds that unqualified aid—particularly aid limited only by costs of attendance—contributes to loan burdens by increasing students' educational costs and their need to borrow (Cellini and Goldin 2014; Lucca, Nadauld, and Shen 2015; Turner 2014). Ultimately, the value of the loan program to students and to the economy at large is determined by the quality and the economic return of the programs it supports relative to the costs of providing that education. This suggests that policymakers should look beyond indicators like default rates and attempt to measure and prioritize value and quality.

Admittedly, that is no easy task. However, improvements in data quality and availability may provide many opportunities to answer key questions, such as how labor market returns differ by program and institution, or how changes in financial aid or public funding affect tuition costs and borrowing. Answers to such questions could help improve the provision of federal aid in the future.

ACKNOWLEDGMENTS The authors thank Sandy Baum, Nick Bloom, Matt Chingos, Natalie Cox, Michael Dinerstein, Naomi Feldman, Daniel Goldenberg, Sarena Goodman, Kerry Grannis, Caroline Hoxby, Adam Isen, Larry Katz, Jordan Matsudaira, Karen Pence, Luigi Pistaferri, Louise Sheiner, Nick Turner, Sarah Turner, David Wessel, Danny Yagan, Owen Zidar, and Eric Zwick, as well as the editors, for helpful comments and discussions. We are also thankful for numerous helpful comments from participants at the Fall 2015 Brookings Panel on Economic Activity. Constantine Yannelis is grateful for financial support from the Alexander S. Onassis Foundation, the Kapnick Foundation, and the Stanford Institute for Economic Policy Research. Any views or interpretations expressed in this paper are those of the authors and do not necessarily reflect the views of the Department of the Treasury or any other organization.

References

Akers, Beth, and Matthew M. Chingos. 2014. "Is a Student Loan Crisis on the Horizon?" Research Report. Washington: Brookings Institution, Brown Center for Education Policy.

Avery, Christopher, and Sarah Turner. 2012. "Student Loans: Do College Students Borrow Too Much—Or Not Enough?" *Journal of Economic Perspectives* 26, no. 1: 165–92.

Barr, Andrew, and Sarah Turner. 2015a. "Out of Work and into School: Labor Market Policies and College Enrollment during the Great Recession." *Journal of Public Economics* 124: 63–73.

———. 2015b. "Aid and Encouragement: Does a Letter Increase Enrollment among UI Recipients?" Working Paper no. 40. Charlottesville: University of Virginia, EdPolicyWorks.

Barron's Educational Series. 2008. "Barron's Profiles of American Colleges." Great Neck and Woodbury, N.Y.

Baum, Sandy. 2015. "Does Increasing Reliance on Student Debt Explain Declines in Entrepreneurial Activity?" Research Report. Washington: Urban Institute.

Bennett, Daniel L., Adam R. Lucchesi, and Richard K. Vedder. 2010. "For-Profit Higher Education: Growth, Innovation and Regulation." Policy Paper. Washington: Center for College Affordability and Productivity.

Bleemer, Zachary, Meta Brown, Donghoon Lee, and Wilbert van Der Klaauw. 2014. "Debt, Jobs, or Housing: What's Keeping Millennials at Home?" Staff Report no. 700. Federal Reserve Bank of New York.

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Variables." *Journal of Human Resources* 8, no. 4: 436–55.

Bound, John, Michael F. Lovenheim, and Sarah Turner. 2010. "Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources." *American Economic Journal: Applied Economics* 2, no. 3: 129–57.

Cellini, Stephanie Riegg. 2009. "Crowded Colleges and College Crowd-Out: The Impact of Public Subsidies on the Two-Year College Market." *American Economic Journal: Economic Policy* 1, no. 2: 1–30.

Cellini, Stephanie Riegg, and Claudia Goldin. 2014. "Does Federal Student Aid Raise Tuition? New Evidence on For-Profit Colleges." *American Economic Journal: Economic Policy* 6, no. 4: 174–206.

Chapman, Bruce. 1997. "Conceptual Issues and the Australian Experience with Income Contingent Charges for Higher Education." *Economic Journal* 107, no. 422: 738–51.

———. 2006. "Income Contingent Loans for Higher Education: International Reform." In *Handbook of the Economics of Education* 2, edited by Eric Hanushek and Finis Welch. Amsterdam: North-Holland.

College Board. 2014. "Trends in Student Aid: 2014." New York.

Deming, David J., Claudia Goldin, and Lawrence F. Katz. 2012. "The For-Profit Postsecondary School Sector: Nimble Critters or Agile Predators?" *Journal of Economic Perspectives* 26, no. 1: 139–64.

Department of Education. 2014. "Student Loans Overview: Fiscal Year 2014 Budget Proposal." Washington: U.S. Department of Education.

Downs, David H., and Pisun (Tracy) Xu. 2015. "Commercial Real Estate, Distress and Financial Resolution: Portfolio Lending versus Securitization." *Journal of Real Estate Finance and Economics* 51, no. 2: 254–87.

Dynarski, Susan, and Daniel Kreisman. 2013. "Loans for Educational Opportunity: Making Borrowing Work for Today's Students." Discussion Paper no. 2013-05. Washington: Brookings Institution, Hamilton Project.

Field, Erica. 2009. "Educational Debt Burden and Career Choice: Evidence from a Financial Aid Experiment at NYU Law School." *American Economic Journal: Applied Economics* 1, no. 1: 1–21.

Gicheva, Dora. 2013. "In Debt and Alone? The Causal Link between Student Loans and Marriage." http://www.uncg.edu/bae/people/gicheva/MBA_loans_marriageJune13.pdf

Gicheva, Dora, and Jeffrey Thompson. 2014. "The Effects of Student Loans on Long-Term Household Financial Stability." Department of Economics Working Paper no. 14-02. Greensboro: University of North Carolina.

Greenstone, Michael, and Adam Looney. 2011. "Where Is the Best Place to Invest $102,000—In Stocks, Bonds or a College Degree?" Online paper. Washington: Brookings Institution, Hamilton Project.

Hartman, Robert W. 1972. "Equity Implications of State Tuition Policy and Student Loans." *Journal of Political Economy* 80, no. 3: 142–71.

Ionescu, Felicia. 2009. "The Federal Student Loan Program: Quantitative Implications for College Enrollment and Default Rates." *Review of Economic Dynamics* 12, no. 1: 205–31.

———. 2011. "Risky Human Capital and Alternative Bankruptcy Regimes for Student Loans." *Journal of Human Capital* 5, no. 2: 153–206.

Johnson, Lyndon B. 1966. "Remarks after a Meeting with Representatives of the American Bankers Association in Connection with the Student Loan Program." Washington, June 10. (Available through the American Presidency Project at http://www.presidency.ucsb.edu/ws/?pid=27643)

Johnson, William. 2006. "Are Public Subsidies to Higher Education Regressive?" *Education Finance and Policy* 1, no. 3: 288–315.

Knapp, Laura Greene, and Terry G. Seaks. 1992. "An Analysis of the Probability of Default on Federally Guaranteed Student Loans." *Review of Economics and Statistics* 74, no. 3: 404–11.

Lucas, Deborah, and Damien Moore. 2010. "Guaranteed versus Direct Lending: The Case of Student Loans." In *Measuring and Managing Federal Financial Risk,* edited by Deborah Lucas. University of Chicago Press.

Lucca, David O., Taylor Nadauld, and Karen Shen. 2015. "Credit Supply and the Rise in College Tuition: Evidence from the Expansion of Federal Student Aid Programs." Staff Report no. 733. Federal Reserve Bank of New York.

Marx, Benjamin M., and Lesley J. Turner. 2015. "Borrowing Trouble: Student Loans, the Cost of Borrowing, and Implications for the Effectiveness of Need-Based Grant Aid." Working Paper no. 20850. Cambridge, Mass.: National Bureau of Economic Research.

Mezza, Alvaro, Kamila Sommer, and Shane Sherlund. 2014. "Student Loans and Homeownership Trends." FEDS Notes. Washington: Board of Governors of the Federal Reserve System.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14, no. 3: 693–709.

Psacharopoulos, George, and George Papakonstantinou. 2005. "The Real University Cost in a 'Free' Higher Education Country." *Economics of Education Review* 24, no. 1: 103–8.

Shao, Ling. 2014. "Debt, Marriage, and Children: The Effect of Student Loans on Marriage and Fertility." Job Market Paper. http://iacs5.ucsd.edu/~lshao/pdfs/shao_jmp.pdf

Sun, Stephen Teng, and Constantine Yannelis. 2016. "Credit Constraints and Demand for Higher Education: Evidence from Financial Deregulation." *Review of Economics and Statistics* 98, no. 1: 12–24.

Turner, Lesley J. 2014. "The Road to Pell Is Paved with Good Intentions: The Economic Incidence of Federal Student Grant Aid." Working Paper. http://econweb.umd.edu/~turner/Turner_FedAidIncidence.pdf

Comments and Discussion

COMMENT BY

CAROLINE HOXBY This paper by Adam Looney and Constantine Yannelis decomposes the striking rise in the volume of student loans and the even more dramatic rise in the default rate on such loans. This is a great paper because it shines much-needed light on the crisis in student loans. It fills, fully but succinctly, what had been a tremendous information gap. It is a tour-de-force demonstration of how useful federal agencies' existing data can be when they are analyzed with the goal of informing policymaking. Indeed, the paper already has been and should continue to be game-changing in debates on student loans. With one small exception, I have nothing but praise for this paper in terms of the questions it asks, how it answers them, and the authors' choice of which material to emphasize.[1]

The paper's most important finding can be stated simply: The rise in student loan defaults is not a puzzle. Rather, the defaults are predictable even with only a limited number of variables that are already observed or readily observable by private lenders and the federal government, including the degree program in which the student is enrolling, the control of the institution (for-profit, nonprofit, public), the program's selectivity, whether the student is nontraditional, and so on. To see why this finding about predictability matters so much, we must first review some economic logic. Afterwards, I return to the implications of the evidence.

GOOD POLICY SOLUTIONS ARE THE RESULT OF ACCURATE DIAGNOSIS Only when a problem has been properly diagnosed are we likely to propose the

1. The one small exception is that the authors focused on default, paying much less attention to income-based repayment. If, as now seems likely, many people in income-based repayment will ultimately not repay their loans, then income-based repayment is merely a slow method of default, and should be counted as such.

correct policy solution for it. Proper diagnosis is almost impossible in the absence of data analysis and a sound theoretical understanding. Suppose you were to go to a physician with a medical problem, and he did not bother to analyze your symptoms or compare them to data he already had about symptoms and diseases. Worse, suppose that he did not even give you a remedy for common symptoms but rather gave you a remedy based on anecdotes in the media about rare diseases. Even worse, suppose that your disease was a consequence of policies that he himself had facilitated. Worst of all, suppose that he had put those policies in place arbitrarily and without logical foundations.

That is our current student loan situation, except that the federal government is playing the part of the physician. The federal government has been aware of the crisis in student loans for some time yet, until this paper was written, did not allow its administrative data to be used by researchers to diagnose the problem. When federal leaders, including the president of the United States, propose solutions to the loan crisis, they motivate their solutions with anecdotes about students who are extremely nonrepresentative.[2] Because (prior to this paper) the federal government did not analyze its own data well, it failed to grasp that the crisis was deeply connected to policies that it itself—not colleges or students or lenders—had insisted upon. Worst of all, the problematic policies have never had a logical basis in economic reasoning. Should we be surprised that policies with no economic foundations produce unintended negative consequences?

APPLYING ECONOMIC LOGIC TO STUDENT LOANS Let us briefly review the economic logic that justifies social or government intervention in the market for student loans. Loans should be the tool of choice *if and only if* the problem is students' being liquidity constrained from investing optimally in their own human capital. If the problem is not liquidity constraints but that education produces social benefits that exceed private benefits, the appropriate solution is tuition that is subsidized (such as we see at public colleges). If higher education investments in students from disadvantaged backgrounds generate especially high social benefits relative to social costs, the appropriate solution is means-tested subsidies (such as the Pell Grant).

2. See, for instance, President Obama's remarks on Ashley in his "Remarks by the President on Opportunity for All: Making College More Affordable," June 9, 2014 (https://www.whitehouse.gov/photos-and-video/video/2014/06/09/president-obama-speaks-student-loan-debt#transcript). In this and numerous other speeches, the president has referred to his own and the first lady's experiences with student loans. These are experiences from which it is hard to derive policy lessons for the typical borrower or defaulter.

If students from disadvantaged backgrounds have especially high option value from trying higher education (because their secondary schools were less likely to inform them about whether and where they could succeed in higher education), then front-loaded subsidies aimed at disadvantaged students are the appropriate solution. I could provide many other examples that map a market failure to an appropriate solution. The point is that the *only* problem for which loans are the appropriate solution is liquidity constraints.

That being said, economic theory also indicates that we should expect failures in the market for student loans. The first and most important problem is that, owing to the illegality of indentured servitude, human capital cannot be put up as collateral for a loan. If a person fails to repay his loan although he is capable of it (moral hazard), a private lender cannot force him to work until the loan is repaid. Moreover, because people know that indenture is illegal, they may choose to take out student loans that they expect not to be able to repay (adverse selection). Since collateral is the main remedy for asymmetric information in loan markets (think of how a bank's ability to repossess a house addresses moral hazard and adverse selection), the uncollateralizability of student loans can generate grave market failures. A second problem is that it might be socially optimal to forgive all or part of a loan if a person suffers from a disabling event or takes up nonremunerative but socially beneficial (public service) work. Private lenders would have great difficulty contracting on such contingencies. Their determination of disability would be subject to question and the definition of public service would surely change over time.

Government intervention in student loans is justified if it remedies the aforementioned problems. In particular, the government has much greater ability to garnish earnings and tax refunds than private lenders do. The government can also outlaw the discharge of student loans in bankruptcy. While these two provisions of the student loan program (wage garnishment, lack of discharge in bankruptcy) do not *solve* the collateral problem, they certainly mitigate it. Additionally, the government has unusual powers to determine and enforce definitions of disability and public service. Under its current loan forgiveness and forbearance programs, the government itself repays some or all of a loan, letting disabled students and public servants "off the hook."

We have now reviewed the economic logic that justifies wage garnishment, lack of discharge in bankruptcy, and certain types of loan forgiveness. However, there is no economic logic for the single most important government intervention in the student loan market: the prohibition on actuarially fair underwriting. That is, the federal government insists that

students face interest rates and loan limits that do not reflect their expected ability to repay. Even if we can easily gather data that show that students who enroll in a particular program have an extremely high risk of default, the federal government insists that those students be offered the same terms as students who enroll in a program for which default risk is negligible. The federal government also prohibits taking account of the share of postsecondary investment that students and colleges make from their own funds, as opposed to borrowed funds. This is akin to forbidding lenders from considering down payments when underwriting mortgages.

There is no economic logic for these prohibitions. Perversely, they introduce market failures that need not exist. Low-default-risk students face unduly high interest rates and unduly low loan limits, causing them to underinvest. Students who would avoid high-default-risk programs if the risks were signaled through fair interest rates and limits do not see any such signals. They thus enroll where they would not if we did not ban their receiving information that lenders would freely make available. Institutions that add little or no value have slight incentive to improve because loans to their students are just as attractive as loans to students who attend schools with a high value added. Moreover, as we shall see, the unjustified prohibitions on fair underwriting are deeply connected with the student loan crisis.

Before returning to the paper and its evidence, it is worthwhile addressing two often-heard concerns. First, students may be high-risk human capital investments through no fault of their own. That is, their disadvantaged backgrounds may have caused them to be poorly prepared for college, uninformed about which program would suit them, and so on. Fair underwriting would thus have the average effect of lowering their loan limits and raising their interest rates. (Note that this would only be the average effect. A disadvantaged student who was well prepared and enrolled in a program with low default rates would, under fair underwriting, enjoy more generous limits and lower interest rates than she does now.) But prohibiting fair underwriting is not an appropriate solution to disadvantaged students' tendencies to be underprepared and underinformed. These are not liquidity problems, and the lack of fair underwriting simply aggravates the information problem. These problems would best be addressed by solutions such as means-tested, front-loaded subsidies, free remediation, or a free first "trial" semester or year of schooling.[3]

3. A better solution would be elementary and secondary education such that students who were equally able and motivated would end up being equally prepared for and informed about college, regardless of background. This is not, however, a proximate remedy.

Second, President Obama and others often say that private lenders, before the federal takeover of student loans, took middleman profits but faced no risk because the government guaranteed repayment.[4] Thus, they would have had no incentive to do fair underwriting had they been allowed to do it. This is true, but the fault was entirely due to the federal government's ignoring economic logic. Logic dictates that the federal government should have employed its enforcement powers (for example, wage garnishment and the ability to outlaw the discharge of loans in bankruptcy) to ensure that payments that could be made did get to private lenders—thereby collateralizing the loans to the extent possible. However, the government need not have guaranteed private lenders against default (except in cases where it wanted to offer loan forgiveness). If the government had simply stopped where economic logic told it to stop, private lenders would have had every incentive to underwrite correctly.

LACK OF FAIR UNDERWRITING IS THE PROXIMATE CAUSE OF THE STUDENT LOAN CRISIS All this discussion of underwriting would be irrelevant if researchers were to find, on digging into loan data, that little of the variation in default rates could be explained. After all, if default rates were unpredictable, lenders would offer all students similar terms even if the government allowed fair underwriting. However, Looney and Yannelis demonstrate that default is predictable even if one uses only a few simple variables that already appear in federal and lenders' databases. They show that default is highly concentrated among students (i) who attend for-profit and, to a lesser extent, public 2-year schools, (ii) who enroll in nondegree, certificate-granting, and 2-year programs, (iii) who attend nonselective schools, and (iv) who are nontraditional. The rise in default rates is not mysterious; rather, enrollment in always-default-prone categories has been rising. In contrast, the probability of default among students enrolled in selective 4-year schools remains so small that such students are almost certainly facing excessively high interest rates and excessively low loan limits. The number of students enrolling in selective 4-year programs has also increased only a little. Indeed, if the authors had gone further and showed how much of the today's default rates could be predicted by a school-times-program fixed effect (default history), they could undoubtedly have shown that fair underwriting would differentiate interest rates and limits by *specific* schools and programs, not merely *types* of schools and programs.

4. See, for instance, President Obama's "Remarks by the President on Higher Education," April 24, 2009 (https://www.whitehouse.gov/the-press-office/remarks-president-higher-education).

Notice that the variables considered by Looney and Yannelis are variables that lenders are allowed to use for underwriting under the Equal Credit Opportunity Act.[5] Indeed, the authors' demonstration of predictability is all the more striking because they denied themselves the use of many variables that would be allowable under the Act. For instance, they did not use students' progress once enrolled in postsecondary school, the track record of students' high schools, or measures of students' incoming preparation.

The point is that if the federal government had not prohibited fair underwriting and, rather, had stuck to the interventions justified by economic logic, the bulk of the default crisis would not exist. Loans with fair rates and limits would have steered students away from programs with low value added. Thus, such schools would have had either to improve or close. Far fewer students would now have loans that they cannot repay. Students would have internalized some of the benefits of studying harder and gaining admission to more selective programs. Economists would not now be concerned about the future budget liability that today's student loans represent. All this would have been accomplished smoothly through the price mechanism—with no need for demagogic speeches, far-fetched "sob stories" about student borrowers, and by-fiat closures of institutions like Corinthian.[6]

(Notice that I said that the *bulk* of the student loan crisis would not have occurred. Looney and Yannelis demonstrate that part of the crisis was transitory. In a recession, more students enroll in postsecondary education because opportunity costs are low but, since the additional students are disproportionately marginal in suitability, they tend to enroll for only short periods. Thus, their repayment problems show up quickly. This causes a temporary surge in default rates. However, problems like this can be addressed by a combination of fair underwriting and conditioning repayment terms on macroeconomic factors. For instance, when unemployment is high among recent college graduates, repayment periods could be extended or back-loaded, keeping each loan's net present value the same.)

What does the Looney and Yannelis paper imply for student loan policy? It does not imply something crude like a ban on loans to students

5. The Equal Credit Opportunity Act (15 U.S.C. 1691) prohibits lenders from discriminating on the basis of race, color, religion, national origin, sex, marital status, age, because an applicant receives income from a public assistance program, or because an applicant has in good faith exercised any right under the Consumer Credit Protection Act.

6. See, for instance, Stratford (2015).

at for-profit schools. Rather, it reminds us to apply economic logic so that society benefits from government interventions that can improve the financing of higher education (grants and tax benefits, as well as loans), but does not suffer from interventions that generate problems that need never have existed.

REFERENCE FOR THE HOXBY COMMENT

Stratford, Michael. 2015. "Corinthian Dismantling Continues." *Inside Higher Ed,* April 15.

COMMENT BY

KAREN PENCE[1] Outstanding balances on government-guaranteed student loans more than tripled between 2000 and 2014, rising from $310 billion to $1.1 trillion.[2] Default rates on these loans increased from 12 to 21 percent over a comparable period.[3] These dramatic changes have raised concerns about student debt burdens and the implications for young Americans' ability to save money, purchase homes, and achieve their life goals.

Policymakers and researchers who want to understand these trends have been stymied by a lack of data, and Adam Looney and Constantine Yannelis have stepped into this void. In a tremendous act of public service, they have merged the student-loan records and wage histories for more than 4 million borrowers from administrative data sets maintained by the U.S. Department of Education and the Internal Revenue Service. They have provided extensive analyses from these data both in their paper and in supplemental spreadsheets in their online appendix available on the Brookings website.

The headline finding from their research is that much of the rise in defaults stems from an increase in the share of borrowers in repayment

1. I am grateful to Ezra Becker, Sarena Goodman, Simona Hannon, Alice Henriques, Felicia Ionescu, Alvaro Mezza, Kamila Sommer, and Chris Smith for help with this discussion. The views in this discussion are mine alone and do not necessarily represent those of the Board of Governors of the Federal Reserve System or its staff.

2. All statistics are from the paper by Looney and Yannelis in this volume or their supplemental online appendix available on the Brookings website unless otherwise noted.

3. The default statistics compare the 3-year cumulative default rates for borrowers in the 2000 and 2011 cohorts as calculated under the Looney and Yannelis preferred method.

who attended for-profit schools or 2-year colleges. This increase stems from two factors. First, as the labor market deteriorated, the number of students who enrolled in college, and took out loans to do so, surged. Between 2008 and 2009, for example, the number of new first-time borrowers increased by 22 percent, and the increases at for-profit and 2-year schools were even higher, at around 32 percent. These students also entered repayment sooner than their counterparts at more selective schools because they were more likely to enroll in short-duration programs or to drop out of school before completing their degrees. For example, between 2010 and 2011, the increase in the number of borrowers entering repayment was 26 percent for for-profit schools, 43 percent for 2-year colleges, and only 13 percent for selective schools.

Borrowers from for-profit and 2-year schools have always had higher default rates than borrowers from other schools. For example, of borrowers who entered repayment in 2011, around 30 percent of those who attended for-profit or 2-year schools were in default three years later, compared with 7 percent of those who attended selective schools. As borrowers from for-profit and 2-year schools became a larger share of all borrowers in repayment, the aggregate default rate mechanically increased. Over time, though, these borrowers will become a smaller share of students in repayment, and the aggregate default rate should decline.

This compositional explanation for the rise in defaults is somewhat unsatisfying, however, as it sidesteps some larger questions. Why are default rates so high for borrowers at these schools? Do these high default rates suggest that the decision to enroll in these schools was a mistake? I will explore these questions in the remainder of this discussion.

To frame these larger questions, it is useful to think of student loan default as the end result of three decisions: to attend school, to finance education with student loans, and to default on those loans. The attendance decision depends on the expected increase in wages after attending school relative to the cost of attendance. The financing decision depends on the cost of student loans relative to other forms of finance. The decision to default depends on the student's ability and willingness to repay the loan.

The attendance decision is complicated because the expected returns to education are uncertain and unfold over many years. The student can affect the returns through the effort that she puts into her studies. But some components of the return—such as future labor market conditions—are outside the student's control, and other factors—such as the quality of the match between the student and the chosen field of study and the idiosyncratic evolution of her life circumstances—may be unknown at the time the

student begins school. A student may make a rational and prudent decision to attend college, and yet still end up with a poor wage realization.[4]

In the aftermath of the 2007–09 recession, job prospects for college graduates appear to have been somewhat worse than after other recent recessions. Job growth after the 2007–09 recession, unlike after the 2001 recession, was disproportionately concentrated in low-wage industries (National Employment Law Project 2014). Likewise, in the aftermath of the 2007–09 recession, the share of college graduates who ended up in jobs that did not require a college degree was somewhat higher than after the previous two recessions (Abel, Dietz, and Su 2014). These findings suggest that young adults could have had reasonable forecasts of the returns to college education—at least if those forecasts were based on the previous two recessions—when they made their enrollment decisions, and in this sense their decision to enroll was not a mistake even if they subsequently struggled to obtain jobs.

The cost component of the attendance decision (tuition and fees) is easier to forecast than the future wages component, although students who assumed that tuition would stay constant might have been surprised by the sharp increases seen over the past three decades. Published tuition prices are now 3.2 times more expensive in inflation-adjusted terms than in 1985 at public universities and 2.4 times more expensive at private universities (College Board 2015, figure 6). If the increase in tuition corresponded to an increase in educational quality, such that expected future earnings were higher, the rise in tuition would be less of a concern. However, at least some of the increase may stem from students bearing more of the cost of their education rather than from an increase in educational quality. At public schools, state budgetary pressures, often coupled with complex state legislative processes, appear to be responsible for some of the tuition rises.[5] At for-profit schools, tuition appears to be sensitive to increases in the government loan limits (Cellini and Goldin 2014). Increases in tuition that do not correspond to increases in educational quality raise the probability of default; students must borrow more to cover their educational expenses, resulting in an increase in debt payments without any increase in expected wages.

4. That said, many students do not make accurate forecasts of the returns to college education. For example, students' enrollment decisions are more sensitive to small changes to tuition and to relatively minor information interventions than would seem optimal if students realized the implications for their lifetime wages (Cohodes and Goodman 2014; Hoxby and Turner 2015).

5. See Goodman and Henriques (2015) for a discussion of the decrease in state support for higher education in the 2000s.

Debt payments also rose during this period because students' incentives increased to finance their educational expenses with government-guaranteed student loans rather than from other resources. First, terms on government-guaranteed loans became more generous during this period. In 2006, for example, the PLUS program, which was formerly open only to parents of students, was expanded to let graduate students borrow as much as the full amount of their tuition and fees; in 2008, aggregate Stafford loan limits increased. Meanwhile, interest rates on Subsidized Stafford loans decreased annually from 2009 to 2012. Second, other sources of funding that might have financed students' education in the past, such as the earnings or home equity of their parents, came under considerable strain during the recession.

Against this backdrop, the fact that defaults were higher in 2014 than in 2000 is not surprising. The actual returns to schooling may have been lower than students expected, whereas debt obligations were higher, meaning that more students would not have the resources to repay their debt. The signs of these strains are apparent even among borrowers who attended selective schools. For example, the share of such borrowers who were unemployed two years after entering repayment rose from 6 percent for the 2000 repayment cohort to 8 percent for the 2012 repayment cohort, and the share whose balances were larger two years after beginning repayment—due to forbearance, repayment plans that allowed payments smaller than the interest, or nonpayment—rose from 25 percent for the 2000 cohort to 36 percent for the 2012 cohort. Meanwhile, the share who defaulted on their student loans within five years of entering repayment rose from 8 percent for the 2000 cohort to 10 percent for the 2010 cohort.[6] For the most part, however, college attendance appears to have been a good investment for these students, with median incomes just below $50,000 in the year that students entered repayment even during the worst years of the recession.

For students who attended for-profit colleges, the outcomes are much worse, almost catastrophically so.[7] Looking at the 2012 repayment cohort, two years after beginning repayment 20 percent were unemployed, 37 percent had incomes at or below the poverty line, and 74 percent had student loan balances that were larger than when they began repayment. Median incomes in the year that students began repayment hovered around $16,500 for all the recession and postrecession repayment cohorts.

6. These increases are more dramatic if measured relative to the experiences of the 2005 cohort: 5 percent of these borrowers defaulted within five years of beginning repayment.

7. A considerable share of borrowers who attended 2-year and nonselective 4-year colleges also had poor outcomes upon entering repayment, although not to the same extent as borrowers from for-profit schools. I focus on for-profit schools here for simplicity.

Why do so many of these borrowers have such terrible outcomes? Three possible and nonexclusive reasons come to mind. First, these borrowers come from families with fewer resources than other students, so they would have struggled more in any circumstance. Indeed, as shown in the paper, borrowers at for-profit schools are disproportionately likely to come from families with few financial resources and around 55 percent of them are first-generation college students.

Second, the education for-profit schools provide may not be effective in increasing borrowers' future earnings. Although this paper does not speak to that question, other researchers have documented that students with credentials from for-profit schools are less likely to be invited for a job interview than comparable students with credentials from public schools (Darolia and others 2015; Deming and others forthcoming). Likewise, students who attend for-profit schools appear to have lower wage trajectories after graduation than other students (Cellini and Chaudhary 2014).

Third, the outcomes for these students may be particularly sensitive to the business cycle. The evidence in the paper supports this idea. For instance, compare students who entered repayment in 2004—who experienced strong labor market conditions for the next two years—to those who entered repayment in 2009, at the depths of the recession. Two years after entering repayment, the unemployment rates for students who attended selective schools were 6 percent for the 2004 repayment cohort and 8 percent for the 2009 cohort; the corresponding shares of borrowers with incomes below the poverty line were 8 percent and 11 percent. For students who attended for-profit schools, the unemployment rates were 14 percent for the 2004 cohort and 21 percent for the 2009 cohort. The equivalent shares with incomes below the poverty line were 26 percent for the 2004 cohort and 35 percent for the 2009 cohort.

What do these data suggest about whether education is a good investment for students who enrolled in for-profit schools? The answer is hard to establish without knowing the wage profiles that the students would have experienced without attending for-profit colleges. However, as one rough benchmark, Jaison Abel and Richard Dietz (2014) suggest that the average income for 25- to 30-year-olds in 2013 with a high school degree was between $25,000 to $30,000.[8] The estimates by Looney and Yannelis indicate that average income for a borrower from a for-profit school during the first couple of years of repayment in 2013 and 2014 was $22,000 to $27,000. This wage comparison—which admittedly does not control for

8. See chart 2 in Abel and Dietz (2014, p. 4).

Table 1. Distribution of Credit Scores in Year before Entering Repayment by Type of School Attended

	Credit Score[a]				
School type	*Less than 550*	*550–600*	*600–660*	*660 or higher*	*No. of observations*
Public 4-year	24	10	17	50	2,879
Private 4-year, nonprofit	20	9	17	54	1,425
Public 2-year	48	11	14	27	1,220
Public, for-profit	63	8	9	20	532

Source: Tabulation by Alvaro Mezza, based on data set described in Mezza and Sommer (2015).

a. Credit score is the TransUnion Account Management Score version 2.0 as measured in the year before the student began repaying the loan. All values are in percentages. The sample spans the years 1998 to 2005.

many important differences between these groups—suggests that attending a for-profit school does not pay off for many students.

What do these data suggest about whether these students are in appropriate debt contracts? Student loan debt, unlike mortgages or auto loans, is not collateralized by objects that a lender can repossess in the event of default. Instead, the collateral is the borrower's future wages. Typically, debt secured by future wages—such as credit card debt—is considered uncollateralized. However, the Department of Education's extraordinary collection authorities, including its ability to garnish borrowers' wages, tax refunds, and Social Security benefits, and the fact that student loans are very difficult to discharge in bankruptcy, have essentially turned student loans into collateralized debt. And indeed, the Department of Education generally recovers at least 80 percent of defaulted loan amounts on a net-present-value basis, after taking collection costs into account (Department of Education, 2014, p. S-31).

Without this ability to attach wages, a lending market might not exist for students at for-profit colleges, because many are not good credit prospects. To illustrate this point, my table 1 shows the distribution of credit scores by type of school attended. The score is measured in the year before students start repayment. In other words, this score measures a student's ability to repay before she has the additional burden of repaying her student loans. The borrowers in this sample started repaying their loans in the 1998–2005 period.[9]

9. The credit score shown is the TransUnion Account Management Score version 2.0. For more details on this sample, see Mezza and Sommer (2015). I thank Alvaro Mezza for creating these estimates.

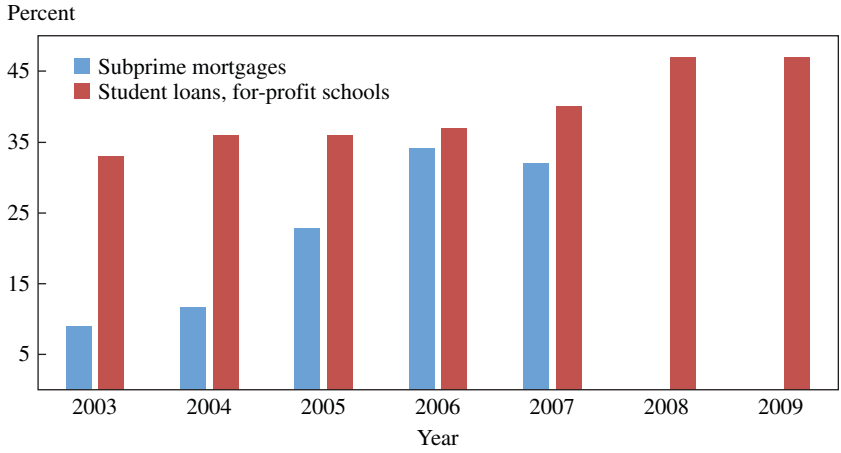
The table indicates that more than 60 percent of students enrolled at for-profit schools had credit scores lower than 550 in the year before they started repaying their loans. To put this score in perspective, borrowers in this category have at least a 12 percent chance of becoming 60 or more days delinquent on any loan in a two-year period.[10] Only 20 percent have a credit score exceeding 660, corresponding to a 4 percent or less chance of becoming 60 or more days delinquent. In contrast, nearly the reverse is true for students who attended a nonprofit private institution before beginning repayment: 20 percent had credit scores less than 550, and 54 percent had credit scores exceeding 660. Alvaro Mezza and Kamila Sommer (2015) show that including these credit scores—again, which are measured before the student enters repayment—substantially improves the predictive power of a model of student loan default.

While the ability to recover collateral may mean that a loan program exists for these students, it also means that students bear all the downside risk of their educational investments. If the returns to education were within students' control, this allocation of risk might be appropriate. However, as noted earlier, returns to for-profit education appear to vary significantly with overall labor market conditions, which are outside students' control. In addition, the fact that more than half of these borrowers are first-generation college students suggests that they may not have much expertise within their networks of family and friends to draw upon in order to evaluate whether attending certain schools is a good investment.

As a thought experiment, it is interesting to compare the default outcomes of borrowers who took out subprime mortgages compared with those who took out student loans to attend for-profit colleges. Both types of borrowers tend to have poorer-quality credit records, and the returns to their investments were dependent on macroeconomic factors beyond their control—house prices in the case of subprime mortgages and wage growth in the case of student loans. At the peak of the housing bubble, both types of loans were about equally risky investments, at least as measured by default rates: Around 35 percent of borrowers who entered into repayment on these products in 2006 defaulted within the next five years (see my figure 1). The total number of borrowers in repayment was also about the same—around 6 million for both loan products in 2006 (my figure 2). The dollars in repayment, of course, were vastly different, as the average dollar

10. The source for this information is the TransUnion Account Management Score, version 2.0, validation odds summary.

Figure 1. Five-Year Cumulative Default Rates by Year Borrower Entered Repayment, 2003–09



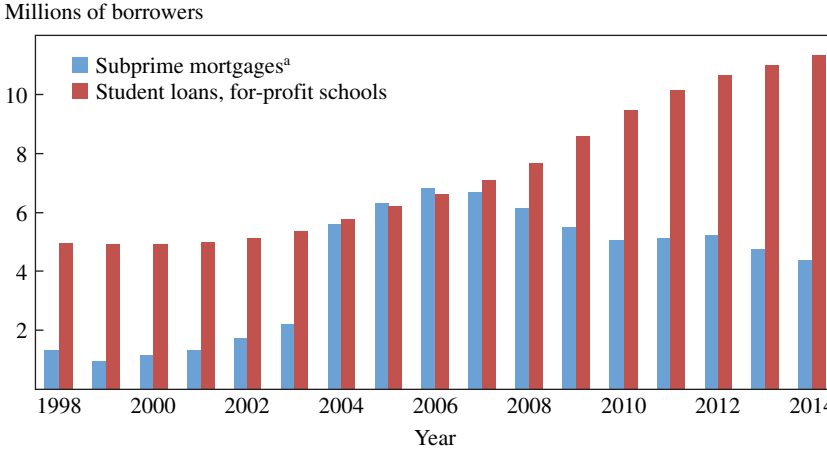
Sources: Looney and Yannelis; Palmer (2015).

balance was around $175,000 for a subprime mortgage and $20,000 for a student loan for a for-profit college (my figure 3).[11]

However, the incidence of the losses was quite different for the two products. Borrowers who defaulted on subprime mortgages lost their homes, but for the most part financial institutions bore the loss of the underwater part of the mortgage—that is, the difference between the mortgage amount and the house value—as well as any interest or fees that accumulated between the time of default and foreclosure, and the expenses associated with maintenance of the property. Borrowers who defaulted on student loans kept the human capital associated with the education, but were not able to discharge the underwater part—that is, the extent to which the increase in lifetime earnings fell short of the loan balance—and were responsible for the interest and fees that accumulated throughout the life of the loan.

11. Subprime mortgages also had profound consequences for the economy and for financial stability that seem unlikely to be repeated in the case of student loans. In addition to being much larger in aggregate dollar volume than student loans originated to borrowers at for-profit schools, subprime mortgages were repackaged into securities that were dispersed throughout the financial system as collateral for a wide variety of financial transactions and obligations. In contrast, the majority of outstanding student loans are held directly by the U.S. government. Subprime mortgages also appear to have had feedback effects on house prices; student loans do not have a similar effect on their collateral (human capital).

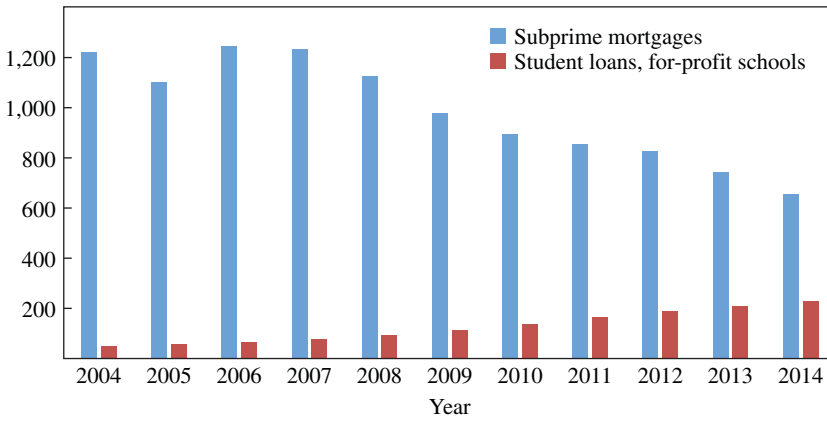
Figure 2. Borrowers with Outstanding Loans, 1998–2014



Sources: Looney and Yannelis; Federal Reserve Board staff tabulation based on data from the Mortgage Bankers Association; McDash Analytics LLC, a wholly owned subsidiary of Lender Processing Services; CoreLogic; the Federal Housing Administration; and the Federal Reserve Statistical Release Z.1, "Federal Accounts of the United States."

a. Assumes number of subprime mortgages is equivalent to number of borrowers.

Figure 3. Outstanding Balances, 2004–14



Sources: Looney and Yannelis; Federal Reserve Board staff tabulation based on data from the Mortgage Bankers Association; McDash Analytics LLC, a wholly owned subsidiary of Lender Processing Services; CoreLogic; the Federal Housing Administration; and the Federal Reserve Statistical Release Z.1, "Federal Accounts of the United States."

The heavy losses borne by private sector financial institutions resulted in a significant contraction of mortgage credit to subprime borrowers. As shown in my figures 2 and 3, the number and dollar amount of subprime mortgage balances have been decreasing since 2006, reflecting defaults and pay-downs on existing loans coupled with almost no new originations. Meanwhile, the fact that borrowers bear the losses for student loans has allowed the student loan market to continue to expand. In 2014, more than 11 million borrowers were repaying student loans originated to attend for-profit institutions, almost twice the number in 2006. Meanwhile, outstanding balances on these loans tripled in nominal terms over this period, from $76 billion to nearly $230 billion. The other major difference between these two types of loans, of course, is that borrowers who defaulted on their mortgages in 2006 have expunged the debt and moved on with their lives. Borrowers who defaulted on their student loans are likely still repaying that debt.

Is there a better way between these two alternatives—one that preserves access to education and credit for disadvantaged borrowers, but also shields borrowers from some of the consequences of macroeconomic events beyond their control? Some initiatives and programs under way will likely help students better forecast the returns to attending different colleges and provide some insurance against poor labor market outcomes. For example, the Department of Education has launched a College Scorecard website (https://collegescorecard.ed.gov) that contains comprehensive data on the average annual costs, graduation rates, and post-attendance earnings for more than 3,500 colleges and universities; these data will help students estimate the returns to education and put pressure on schools to improve outcomes for their students. Enhancing these data with forecasts of students' debt payments relative to earnings might further improve transparency. The Department of Education has also designed a variety of income-based repayment plans, under which a student's payment fluctuates with her income, and any unpaid balances are forgiven after 20 or 25 years. However, some students will end up paying more interest under these plans, and may owe taxes on any forgiven debt balance.

One possibility is giving schools more of an equity stake in their students' success or failure. Schools are in a better position than lenders to monitor the students that they are admitting, and can influence the quality of the students' education and thus their eventual economic outcomes. Under existing regulations, schools lose their eligibility to participate in the Pell Grant and Direct Loan programs if their default rates exceed certain thresholds. Additional proposals along these lines might better align incentives and

improve outcomes for students. However, because schools have the option to go out of business, such proposals will not be able to align incentives fully.

REFERENCES FOR THE PENCE COMMENT

Abel, Jaison R., and Richard Dietz. 2014. "Do the Benefits of College Still Outweigh the Costs?" *Current Issues in Economics and Finance* 20, no. 3.

Abel, Jaison R., Richard Dietz, and Yaqin Su. 2014. "Are Recent College Graduates Finding Good Jobs?" *Current Issues in Economics and Finance* 20, no. 1.

Cellini, Stephanie Riegg, and Latika Chaudhary. 2014. "The Labor Market Returns to a For-Profit College Education." *Economics of Education Review* 43: 125–40.

Cellini, Stephanie Riegg, and Claudia Goldin. 2014. "Does Federal Student Aid Raise Tuition? New Evidence on For-Profit Colleges." *American Economic Journal: Economic Policy* 6, no. 4: 174–206.

Cohodes, Sarah R., and Joshua S. Goodman. 2014. "Merit Aid, College Quality, and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy." *American Economic Journal: Applied Economics* 6, no. 4: 251–85.

College Board. 2015. "Trends in College Pricing 2015." New York.

Darolia, Rajeev, Cory Koedel, Paco Martorell, Katie Wilson, and Francisco Perez-Arce. 2015. "Do Employers Prefer Workers Who Attended For-Profit Colleges? Evidence from a Field Experiment." *Journal of Policy Analysis and Management* 34, no. 4: 881–903.

Deming, David J., Noam Yuchtman, Amira Abulafi, Claudia Goldin, and Lawrence F. Katz. Forthcoming. "The Value of Postsecondary Credentials in the Labor Market: An Experimental Study." *American Economic Review.*

Department of Education. 2014. "Student Loans Overview: Fiscal Year 2015 Budget Proposal." Washington. http://www2.ed.gov/about/overview/budget/budget15/justifications/s-loansoverview.pdf

Goodman, Sarena F., and Alice M. Henriques. 2015. "Attendance Spillovers between Public and For-Profit Colleges: Evidence from Statewide Changes in Appropriations for Higher Education." Finance and Economics Discussion Series no. 2015-025. Washington: Board of Governors of the Federal Reserve System.

Hoxby, Caroline M., and Sarah Turner. 2015. "What High-Achieving Low-Income Students Know about College." *American Economic Review* 105, no. 5: 514–7.

Mezza, Alvaro A., and Kamila Sommer. 2015. "A Trillion Dollar Question: What Predicts Student Loan Delinquencies?" Finance and Economics Discussion Series no. 2015-098. Washington: Board of Governors of the Federal Reserve System.

National Employment Law Project. 2014. "The Low-Wage Recovery: Industry Employment and Wages Four Years into the Recovery." Data Brief. New York.

Palmer, Christopher. 2015. "Why Did So Many Subprime Borrowers Default during the Crisis: Loose Credit or Plummeting Prices?" Job Market Paper. http://faculty.haas.berkeley.edu/palmer/cpalmer-subprime.pdf

GENERAL DISCUSSION Much of the discussion centered around whether it was the characteristics of the *students* or the *schools* that were more important in the apparent student loan debt crisis. David Romer observed that a common narrative is that the problems associated with rising student debt and default are attributable to the rise in for-profit schools, but that an alternative story is that the students who were more likely to have trouble in college began going to for-profit schools and community college when previously they were largely not going to school at all. He noted that Adam Looney and Constantine Yannelis's decomposition analysis appeared to provide significant support for the alternative story. His reading of their results was that they indicated that a smaller proportion of the rise in default rates was attributable to shifts in the composition of schools, and that a larger proportion came from student characteristics. Students who were likely to have difficulty in school or difficulty repaying student loans were going disproportionately to for-profit schools and community colleges. Romer therefore suggested that more emphasis be put on the students, and less on the schools themselves.

Martin Feldstein engaged discussant Caroline Hoxby regarding a claim that roughly two-thirds of student loan defaults are predictable. In response, Hoxby noted that the U.S. Department of Education, through its longitudinal surveys, can match very basic characteristics of the student—such as age and gender—with basic information on the student's achievement—such as test scores, high school grades, and the high schools they come from—to predict roughly 75 percent of the variation in default and repayment problems. She also agreed with Romer that the characteristics of the students are more important in predicting default than the institutions they attend, though the institutions do matter some. At the end of the day, it is the students who are good risks or bad risks, and the institutions add to that.

Henry Aaron stressed that even if default was more strongly associated with the characteristics of students than of schools, this did not imply that the students were more to blame than the institutions, a view he called "exceedingly misleading." On the contrary, he argued that public policy flawed in design, as Hoxby emphasized, had created an enormous financial incentive for private institutions to exploit a particular population. To then say that the students are to blame for the resulting exploitation places responsibility in the wrong place, and is a misdirection in terms of appraising where the remedies to the policy come from.

Christopher Carroll also thought that accountability needed to lie with the institutions and less so with the students. While there is a lot of idio-

syncratic risk of default at the level of the individual person, the responsibility is at the place where the insurance across persons happens, namely, at the institution. He suggested that if an institution's students historically have a high rate of default on their loans, then the institution should have to put up collateral to the federal government, which would effectively accomplish the sharing of the idiosyncratic risk in a much more efficient way than is available to the students themselves. The right focus is the schools, he concluded; in order to align incentives properly while still allowing for insurance, it is the schools that should be held accountable.

Bruce Fallick agreed with Aaron and Carroll that more emphasis should be given to the institutions. One objective of public policy in the education domain, he claimed, is to allow students from disadvantaged backgrounds, who may have the capacity, to have the opportunity to achieve higher education. Traditional nonprofit institutions have an incentive to evaluate students on the basis of information beyond what is readily available in the public domain. By cutting off loans by institutional characteristics, as opposed to by individual characteristics, these incentives can be more easily aligned, and the "diamond in the rough" can more easily obtain higher education, he concluded.

Alan Blinder noted that one reason why student loans involve "horrendous" underwriting is that the distribution of income inevitably excludes some students who otherwise would not be able to avail themselves of higher education. By and large, it is a good idea to help those people, according to Blinder. He wondered if, in terms of policy, this might push the conversation away from loans and more toward grants.

Frederic Mishkin emphasized that what is important is thinking about exactly what the market failure is when it comes to student loans, which Douglas Elliott seconded. Perhaps it is an information problem, Mishkin suggested, in which students are not getting the information that they need to make informed decisions about higher education. The idea that for-profit institutions are just "bad guys" who need to be taken care of is not the right way to think about solving the problem, he argued. Focusing on the market failure aspect may help bring to light the kinds of innovations that could come from for-profit institutions, particularly in the online sector. The idea of simply closing down or otherwise severely punishing for-profit institutions, he concluded, could actually be very bad public policy.

Greg Mankiw asked the authors and discussants to comment on two ideas that have been floated that are related to postsecondary education and student debt. The first, most notably advocated by Charles Murray, is the idea that too many people are being pushed into postsecondary

education. Murray has argued that some people are not cognitively ready for college, and that more people should be pushed into apprenticeships after high school.

The second, which Mankiw associated with Luigi Zingales, is the idea of having equity investment in students. That is, a private investor could pay for a student's education in return for a small percentage of her eventual income over, say, the next 30 years. He added that the Internal Revenue Service might be the best agency for enforcing such a contract because it actively observes incomes. If such a policy were ever implemented, investors would have incentives to pick the best students and direct them to the best programs. Such a system, as opposed to the current student loan program, might provide higher incentive compatibility between the student and the financier. Aaron agreed that it might be interesting to play with the incentive schemes that might emerge under such a contingent repayment arrangement.

Jan Eberly suggested that future work look more seriously at PLUS loans. She noted that PLUS loans, which are generally held by parents of the students, do not have the same flexibility as federal undergraduate loans, and often they are not only held by parents, but sometimes by grandparents, aunts, uncles, and other adults who may have very little influence on the behavior of the borrower. PLUS loans, she concluded, are large and not well understood.

Adam Looney conceded that a lot of the increase in default rates among for-profit borrowers is unexplained by regression analysis. The real question, he argued, is how to apportion between the characteristics that are observed in the data and the unobserved characteristics of the students or the institutions. One explanation for why students sometimes default is that they were unsatisfied with their education when they left, which is generally not observable in the data, he noted.

Regarding Blinder's comments about grants and the distribution of income, Constantine Yannelis noted that student loans under the current system are not an effective form of redistribution, primarily for two reasons. First, student loans are not dischargeable in bankruptcy, and wages as well as Social Security payments can be garnished later in life. Secondly, student loans are regressive in that people who borrow more tend to earn more in the future.

On the question of grants, Yannelis noted that from the free market perspective, there is something very attractive about grants and vouchers more generally, and he wondered why they seem to perform so poorly in the education space. The current structure of the education market is one

in which nobody is paying with their own money, and where vouchers have virtually zero shadow cost. There is a certain market tension missing, and there does not seem to be a real cost to the investor. That tension is in play in a place where nobody is playing with their own money, and it is also in play when there are no loans and only grants.

Regarding the problem of underwriting, Yannelis noted that structuring the right kind of underwriting is a tough thing to think about because one would not want it to be conditioned on, for example, a student's economic background. Furthermore, the individuals themselves are probably not in a great place to understand what institution is going to be right for them, and whether they are qualified to make that decision. The institution, he argued, acts as an intermediary; by focusing more on the institutions than on the individuals as the place where the incentives could be strengthened, that could help achieve higher returns. In the current system, incentives are not aligned between the institutions and the students because the institutions do not currently bear the costs of defaults.

Regarding Mankiw's first question about whether or not some people should even go to college, Yannelis expressed optimism that there are higher returns to education more generally, including way down the income distribution. He noted that there are already some postsecondary programs that cater to extremely disadvantaged people, are entirely reliant on aid, and generally produce fairly good outcomes. He observed that it seems like there are way too few people who are taking advantage of those programs, and way too many people who are entering programs that they cannot finish and that do not lead to a degree that produces a high return. He advocated for trying to move people into places where they are going to have a high return, and to move people out of places where they have very low returns.

On Mankiw's second question about human capital contracts, Yannelis noted that they depend crucially on adverse selection into those contracts, and that currently there is not a lot of information about those parameters. He noted that Yale University tried a similar experiment in the 1970s, though the plan effectively fell apart precisely due to adverse selection concerns. On the other hand, some countries like Australia have income-based repayment systems that everyone is defaulted into, and that appears to work fairly well.

DOUGLAS W. ELMENDORF
Brookings Institution[1]

"Dynamic Scoring": Why and How to Include Macroeconomic Effects in Budget Estimates for Legislative Proposals

ABSTRACT Official estimates of the budgetary effects of legislative proposals generally include anticipated behavioral responses except for those that would alter overall output or employment. Based on my experience as director of the Congressional Budget Office and on the analysis in this paper, I conclude that such macroeconomic effects of legislative proposals should be included in budget estimates—that is, so-called dynamic scoring should be used—for major (but not minor) proposals and for proposals affecting federal spending as well as revenues. However, such macroeconomic effects should not be included when the estimating agencies do not have the tools or time needed to do a careful analysis of those effects. Current rules governing the official estimating process do not fully meet those conditions.

When legislation is being developed in the U.S. Congress, the Congressional Budget Office (CBO) and the staff of the Joint Committee on Taxation (JCT) prepare estimates of the effects of that legislation on the federal budget. Those estimates often play a critical role in congressional deliberations and public discussion.

The estimates produced by CBO and JCT generally incorporate the effects of anticipated behavioral responses to the proposed changes in federal tax or spending policies. For example, estimates for changes in benefit programs include shifts in take-up rates among eligible people, and estimates for changes in income tax rates include shifts in the use of tax deductions. However, by long-standing convention, the estimates

1. The author was working at Brookings when he produced this paper; in January 2016 he became dean of Harvard's John F. Kennedy School of Government.

have excluded behavioral responses that would have macroeconomic effects, in the sense of altering overall output, employment, or similar variables. For example, CBO and JCT's original estimate of the budgetary impact of the Affordable Care Act (ACA) included the effects of employers altering the mix of taxable and nontaxable compensation provided to their employees but not the effects of employees altering their supply of labor.

The convention of excluding macroeconomic effects may seem odd from an economics perspective. Estimates for legislative proposals include behavioral responses in order to improve the accuracy of the predicted budgetary effects and to illuminate noteworthy nonbudgetary effects, and that rationale appears to apply equally to behavioral responses that affect overall output and those that do not. Indeed, some analysts and policymakers have argued for years that the estimates produced by CBO and JCT should include macroeconomic effects—an approach that has become known as "dynamic scoring." However, other analysts and policymakers have argued in response that including macroeconomic effects would degrade the quality and usefulness of CBO's and JCT's estimates.[2] That debate has achieved greater prominence recently because a rule adopted by the House of Representatives and the budget resolution approved by the House and the Senate both call for dynamic scoring in certain circumstances.

Based on my experience as the director of CBO from January 2009 through March 2015, I believe the principal concerns expressed about estimating the macroeconomic effects of proposals apply with equal force to other aspects of budget estimates or can be addressed by CBO and JCT. In my view, including macroeconomic effects in budget estimates for certain legislative proposals would improve the accuracy of those estimates and would provide important information about the proposals' economic effects. Moreover, if certain key conditions were satisfied, those estimates would meet the general goals of the estimating process, namely: that estimates be understandable and resistant to misinterpretation, that they be based on a consistent and credible methodology, that they be produced quickly enough

2. The advantages and disadvantages of dynamic scoring have been considered by numerous authors, including Auerbach (1996, 2005), Burman (2006), Committee for a Responsible Federal Budget (2012), CBO (1995, 2002), Furman (2006), Gale (2002), Hassett (2002), Holtz-Eakin and Mandel (2015), Ip (2015), Orszag (2002), and Van de Water and Huang (2014).

to serve the legislative process, and that they be prepared using the resources available to CBO and JCT.

Therefore, I conclude that the macroeconomic effects of legislative proposals should be included in budget estimates—that is, that dynamic scoring should be used—under the following conditions:

—Macroeconomic effects should be included in estimates only for *major* proposals, defined as those that would have a large estimated budgetary impact excluding macroeconomic effects, and when estimates of such effects are requested by the chair or ranking member of the House or Senate Budget Committee. CBO and JCT do not have sufficient staff or time to carefully analyze macroeconomic effects for every proposal under consideration, and using rules of thumb in place of careful analysis risks the credibility of the estimates.

—Macroeconomic effects should be included in estimates for major proposals affecting federal *spending* as well as revenues. Changes in either spending or revenues can have notable macroeconomic effects, and the estimating process should treat proposals affecting the two sides of the budget as comparably as possible subject to other constraints.

—Macroeconomic effects should *not* be included in estimates when CBO and JCT find that they do not have the tools or time needed to do a careful analysis of those effects. That situation will arise most often for proposals that are being developed and amended quickly and for proposals regarding certain types of regulatory policy in which the estimators do not have significant expertise.

Those conditions, and others discussed in this paper, can be readily satisfied. However, the current House rule and congressional budget resolution do not fully meet the specified conditions. The current requirements for dynamic scoring explicitly exclude appropriations bills (which cover about one-third of federal noninterest spending) and give only the chairs but not the ranking members of key committees the right to request the incorporation of macroeconomic effects in certain estimates. In addition, the threshold budgetary impact for presumptively including macroeconomic effects in estimates is lower than ideal, from my perspective.

There are advantages to an alternative approach in which CBO's and JCT's estimates of macroeconomic effects and their budgetary feedback would be provided in supplementary reports rather than being included in official budget estimates. In my judgment, though, the advantages of that alternative approach, as compared with the agencies' current plans for dynamic scoring, are limited and are outweighed by significant disadvantages.

I. The Basics of Budget Estimates for Legislative Proposals

CBO and JCT provide the official estimates used by Congress of the effects of legislative proposals on the federal budget. The estimates are based on procedures that have been developed over time and on the professional judgment of the two agencies. The analysts at CBO and JCT stay in their jobs regardless of political shifts in the control of Congress (although the director of CBO and the chief of staff for JCT are chosen by the congressional leadership), and the organizations have strong reputations for providing objective, nonpartisan analysis.

I.A. The Mechanics of Estimates

CBO, which began work in 1975, produces public estimates for bills after they have been approved by congressional committees or before they are voted on by the full House or Senate. For bills that would alter the tax code, CBO is required by its founding statute to use revenue estimates provided by JCT, which was created in 1926; for bills that would alter spending policies, CBO uses its own estimates; and for bills that would make changes in both tax and spending policies, the agencies prepare estimates together. The estimating process is sometimes referred to as "scoring," and the estimates are called "cost estimates." In addition to those public estimates, the agencies provide private estimates to members of Congress and their staffers for proposals that are being developed and have not been released publicly. In a typical year, CBO publishes between 500 and 600 public estimates, and it and JCT give committees thousands of private estimates for legislation under development.

Each estimate shows effects relative to the "baseline," which is CBO's projection of what would occur in the absence of the proposal. The baseline generally reflects current law, although Congress has specified certain exceptions.

The estimates present changes in nominal cash flows for the current fiscal year and each of the 10 subsequent years, a period that is often called the "budget window."[3] The use of cash flows and a limited time period

3. The principal exception to this statement is estimates for federal credit programs, which are based on the accrual of financial commitments by the federal government (CBO 2012a). All cost estimates exclude changes in federal interest payments that would result from changes in federal borrowing. However, CBO includes changes in interest payments when it provides estimates for overall budget packages, as in its annual analysis of the President's budget proposals.

mean that estimates do not always measure a proposal's full budgetary effect. Indeed, some proposals deliberately delay cash costs beyond the budget window or accelerate cash receipts into the budget window in order to lower the apparent budgetary impact.[4] However, when Congress is especially interested in a proposal's long-term budgetary effects, the agency tries to provide information about those effects. And when CBO expects that a major proposal would have notably different budgetary effects beyond the coming decade than during the decade, the agency can provide information about those effects without a specific request from the Congress.

The estimates are point estimates that are intended to show what is colloquially described as "the middle of the distribution of possible outcomes," but is specifically the mean outcome as judged by the agencies (CBO 1999). Although CBO and JCT are acutely aware of the uncertainty of estimates, the agencies focus on point estimates because the budget process and the procedural rules of the House and Senate rely on point estimates and because measuring the uncertainty of estimates is often especially difficult.

I.B. Behavioral Responses

CBO's and JCT's estimates generally include the impact of behavioral responses to the proposed changes in law—that is, the estimates are not based on an assumption that the economy is static. For example, estimates of changes in benefit programs include shifts in take-up rates for those benefits among eligible people, and estimates for changes in income tax rates include shifts in the use of tax deductions. More generally, CBO and JCT try to account for the behavior of households, businesses, federal regulators, and state, local, and foreign governments. However, the agencies do not attempt to predict future changes in federal law. CBO and JCT estimate the magnitude of behavioral responses using a broad range of evidence, including formal statistical analyses done by the agencies themselves and by other researchers as well as anecdotal information from consultations with government agencies and private businesses (CBO 2011; JCT 2011a).

The scope of the included behavioral responses varies greatly across estimates. Some potential responses are omitted because the available evidence does not indicate the order of magnitude or even the sign of a response.

4. For example, in a policy change known as "pension smoothing," companies are allowed to defer required payments into pension funds, thereby increasing their reported profits and thus tax payments in the budget window while reducing them later.

Other potential responses are omitted because CBO or JCT do not have the time or resources to collect the available evidence and build and apply an appropriate model. Still other potential responses are omitted from estimates because including the responses would greatly complicate the legislative process; for example, the effects that certain changes in spending would have on taxable incomes and thus on revenues are omitted from estimates because including them would generate jurisdictional conflicts between committees.

In addition, some potential behavioral responses are excluded from estimates because the responses would affect overall output, and overall output has been held fixed in cost estimates by long-standing convention. Therefore, CBO's and JCT's estimates have not included the budgetary effects of changes in labor supply, consumption, saving, productivity, and other aggregate variables. Under dynamic scoring, this convention of "fixed output" would be dropped. The principal exception to this convention before 2015 was the production of estimates for comprehensive immigration legislation in 2006, 2007, and 2013. In CBO's view, "assuming that those bills would have had no effect on overall output would have ignored one of the primary effects of the bills and distorted those estimates too severely" (CBO 2015j).

I.C. Current Status of Dynamic Scoring

In early 2003, the House adopted rule XIII.3.(h)(2), which required JCT to provide an analysis of the macroeconomic impact of all tax legislation approved by the Ways and Means Committee. That rule was adopted again by subsequent Congresses and remained in effect in the House through 2014. JCT (2015a, p. 12) summarized its response to this rule as follows:

> For most tax bills, the expected effects were so small that a brief statement to that effect was all that was required. Short qualitative analyses were provided for legislation that JCT macro models were not configured to model. For major tax legislation, JCT staff has provided detailed quantitative analysis of a possible range of effects on GDP, employment, investment, and revenues, based on the results of multiple models using multiple parameter assumptions.

In early 2015, the House modified that rule to require dynamic scoring by CBO and JCT for "major" legislation, defined as legislation that would have significant estimated budgetary effects or was designated as major by the chair of the Budget Committee or the Ways and Means Committee. The rule excludes appropriations bills, requires a qualitative assessment of

budgetary impact including macroeconomic effects for 20 years beyond the 10-year budget window, and includes the caveat that the requirements be met "to the extent practicable."

In the spring of 2015, the House and Senate approved a budget resolution that included requirements for dynamic scoring similar to those in the House rule. Under the budget resolution, CBO and JCT will, to the greatest extent practicable, incorporate the budgetary effects of changes in macroeconomic variables resulting from legislation that has a "gross" budgetary effect of a quarter of a percent of output in any year over the next 10 years, or is selected for such analysis by the chair of the House or Senate Budget Committee. That threshold equals about $45 billion in 2015 and about $70 billion in 2025 based on projected output (CBO 2015a); I address the interpretation of "gross" budgetary effects later in the paper. The resolution excludes appropriations bills and requires a qualitative assessment regarding the two decades following the budget window.

CBO (2015i) summarized the agency's plans for meeting the requirements of this budget resolution. In June 2015, in response to a request from the Senate Budget Committee, CBO and JCT applied dynamic scoring to a proposal to repeal the Affordable Care Act (CBO 2015h). And in August 2015, pursuant to the resolution, JCT applied dynamic scoring to a bill approved by the Senate Finance Committee that would extend for 2 years a number of tax credits, deductions, and exclusions that primarily affect businesses (JCT 2015b; CBO 2015k).

II. CBO's and JCT's Past Estimates of Macroeconomic Effects and Their Budgetary Feedback

The arguments for and against dynamic scoring can be understood best after briefly examining CBO's and JCT's past analysis of proposals' macroeconomic effects and their budgetary feedback. Most of that analysis has appeared not in cost estimates—given the conventional exclusion of macroeconomic effects from such estimates—but rather in supplemental reports.

II.A. Analytic Approach

CBO (2014d) provided an overview of its methodology for estimating the macroeconomic effects of legislative proposals in a November 2014 report, "How CBO Analyzes the Effects of Changes in Fiscal Policies on the Economy"; that report referred to a set of other reports describing specific aspects of the agency's methods (CBO 2001, 2012d, 2012f, 2012g, 2012h, 2013e, and 2014b). JCT (2003b, 2005, 2006, 2011a, 2011b, and

2015a) has summarized its methodology as well. In addition, in each report that presents such a macroeconomic analysis, CBO and JCT explain the key factors affecting the estimates.[5]

The approaches used by the two agencies differ in various specifics but are quite similar in their overall structure. For estimating the *short-term* effects of changes in fiscal policies, the agencies focus on effects on the demand for goods and services (and also include effects from changes in labor supply). Reductions in taxes and increases in federal spending boost demand directly, while the opposite changes diminish it; those direct effects propagate through the economy to an extent that depends on the response of monetary policy and other factors.[6] Changes in demand are estimated to lead to changes in output relative to potential output.

For estimating the *longer-term* effects of changes in fiscal policies, the agencies examine effects on potential output. Both agencies use a Solow-type growth model and a life cycle (overlapping generations) growth model.[7] In its Solow-type model, CBO focuses on the effects of changes in federal borrowing, marginal and average tax rates (through income and substitution effects), transfer payments (through income effects and, in some cases, substitution effects), and federal investment in physical infrastructure, education and training, and research and development. For example, an increase in the marginal tax rate on labor income is estimated to reduce the supply of labor, which in turn reduces capital accumulation. Similarly, JCT's Solow-type model captures responses to changes in federal borrowing, marginal and average tax rates, and other factors. In their life cycle models, CBO and

5. There are recurring calls for CBO and JCT to be more transparent regarding many aspects of their analyses. However, achieving greater transparency would require the agencies to allocate more of their resources to explaining existing estimates rather than producing new ones, and Congress has been reluctant to accept that trade-off.

6. CBO projects actions by the Federal Reserve as part of its baseline economic projections, and the agency has explained its method for estimating the Federal Reserve's reaction to changes in fiscal policies. For example, CBO (2015g) expects that the negative short-term effects of deficit reduction on output (stemming from a decrease in demand) will be "stronger when short-term interest rates are near zero . . . because under those conditions the Federal Reserve is unlikely to adjust short-term interest rates to try to offset the effects of changes in federal spending and taxes" (p. 88).

7. JCT (2011a, 2011b) also sometimes uses a growth model with infinitely lived agents. Separately, CBO (2014d, p. 12) discusses the possibility that changes in demand in the short term could affect potential output in the long term and concludes that the significance of the channels through which that might occur are "unclear," and thus "CBO does not currently incorporate such channels in its analyses, although the agency continues to investigate the issue."

JCT include many of the same channels, although expectations also matter explicitly—which presents a challenge for dynamic scoring that I return to below.

CBO generally reports both a central estimate and a range of estimates for the macroeconomic effects of proposals. The range is based on the values for the key parameters found in the research literature; the range for each variable "is intended to cover roughly the middle two-thirds of the likely values for the variable" (CBO 2015g, p. 73). The central estimate is intended to represent the middle of the distribution of possible outcomes (and can give weight to estimates from both the Solow-type model and the life cycle model).

To estimate the feedback from economic changes to the federal budget, CBO accounts for the impact of changes in income on tax revenues and benefits (with the latter much less affected than the former), as well as other factors. A one-dollar increase in overall output reduces the budget deficit by roughly 20 to 25 cents, holding all else equal.[8] The estimated budgetary effects in a given year influence estimated economic developments in subsequent years.

A key challenge for CBO and JCT is assessing the changes that proposals would generate to effective marginal tax rates on labor and capital, the income of people with different propensities to consume, differences in tax rates across types of capital, and changes in federal investment. CBO and JCT also modify their models as needed to capture the features of specific proposals. That process—as undertaken, for example, in the analyses of immigration reform and tax reform discussed below—sometimes requires a great deal of time and effort. In addition, the agencies adjust parameter values over time in response to new evidence; for example, see CBO's (2012g) paper "A Review of Recent Research on Labor Supply Elasticities."

CBO's and JCT's analyses of the macroeconomic effects of proposals are generally produced on much longer timetables than their budget estimates. That difference arises both because estimating macroeconomic effects can take considerable time and because the estimated budgetary impact (excluding macroeconomic effects) is one of the inputs into estimating a proposal's macroeconomic effects.

8. CBO (2015a, p. 133) provides a rule of thumb for the budgetary impact of lower output growth.

II.B. Comprehensive Immigration Legislation

In 2013, the Senate passed a bill to substantially increase the number of people who could enter the country legally and to create a process through which many people who are currently present in the country on an illegal basis could gain legal status. CBO and JCT provided estimates for the legislation in two separate documents released simultaneously: a cost estimate that included some but not all of the expected macroeconomic effects of the bill (CBO 2013b), and a supplemental analysis of the bill's total macroeconomic effects and the incremental budgetary impact of the economic changes not included in the cost estimate (CBO 2013c). CBO (2013c, p. 2) explained the analysis this way:

> [Since the legislation] would significantly increase the size of the U.S. labor force, assuming that total employment was unchanged would imply that any employment of the additional immigrants would be offset one-for-one by lower employment elsewhere in the population. Because that outcome would be highly implausible, CBO and JCT relaxed the assumption of fixed GDP and employment and incorporated into the cost estimate their projections of the legislation's direct effects on the U.S. population, employment, and taxable compensation. Nevertheless, to remain as consistent as possible with the estimating rules CBO and JCT follow for almost all other legislation, the cost estimate . . . does not incorporate the budgetary impact of every economic consequence of the bill. The [supplemental] analysis . . . includes some additional budgetary effects stemming from changes in the productivity of labor and capital, the income earned by capital, the rate of return on capital (and therefore the interest rates on government debt), and the differences in wages for workers with different skills.

That is, the cost estimate excluded macroeconomic changes that could be excluded without making the estimate nonsensical, and the supplemental analysis included all of the macroeconomic changes that CBO was able to estimate.

In the cost estimate, CBO and JCT estimated that the bill would reduce cumulative budget deficits by about $200 billion during the first decade after enactment and about $700 billion during the following decade. In the supplemental report, CBO estimated that the bill would raise output by roughly 3 percent by the end of the first decade; additionally, it estimated that economic effects not included in the cost estimate would have no further net effect on the cumulative deficit in the first decade but would further reduce the cumulative deficit in the second decade by about $300 billion.

II.C. Affordable Care Act

When CBO and JCT estimated the budgetary effects of the ACA and its precursors in 2009 and 2010, they incorporated the impact of many changes

in the behavior of individuals, employers, health insurers, and health care providers. However, the agencies did not include the impact of certain other changes in behavior because of the long-standing convention for cost estimates that overall output would be unaffected.

By contrast, given the recent congressional push for dynamic scoring, CBO and JCT's estimate in mid-2015 of the effects of *repealing* the ACA included macroeconomic changes. The estimate incorporated, among other factors: short-term effects on aggregate demand of changes in federal spending and taxes; effects on labor supply of changes in tax rates for certain higher-income people and of changes in subsidies for health insurance for certain lower-income people; and effects on capital investment of changes in tax rates, labor supply, and federal borrowing. The largest macroeconomic impact of repealing the ACA was estimated to stem from repealing the subsidies for health insurance.[9] The agencies concluded:

> Repealing the ACA would increase federal budget deficits by $137 billion over the 2016–2025 period . . . , [which incorporates] the net effects of two components: Excluding the effects of macroeconomic feedback . . . , federal deficits would increase by $353 billion over the 2016–2025 period if the ACA was repealed. Repeal of the ACA would raise economic output, mainly by boosting the supply of labor; the resulting increase in GDP is projected to average about 0.7 percent over the 2021–2025 period. Alone, those effects would reduce federal deficits by $216 billion over the 2016–2025 period. (CBO 2015h, p. 1)

II.D. Congressman Camp's Tax Reform Proposal

In early 2014, Congressman Dave Camp, then the chair of the Ways and Means Committee, put forward a comprehensive proposal for broadening the bases of the individual and corporate income taxes, adjusting tax rates,

9. The estimate of the effects on labor supply drew heavily on CBO's published analysis during the preceding several years. CBO (2009c) examined various channels through which changes to the health insurance system could affect labor markets; however, specific proposals were still in formative stages at the time, so the report did not provide quantitative estimates. In late 2009 and early 2010, congressional interest in the evolving health care legislation focused on its effects on the federal budget, health insurance coverage, insurance premiums, and existing federal programs, so CBO and JCT's analysis focused on those issues. After the ACA was enacted, CBO needed to incorporate the law's economic effects into the baseline economic projections. As part of that process, CBO (2010b, pp. 48–49) reported that it expected the ACA to reduce aggregate labor supply by an amount that would reduce labor compensation by roughly one-half percent after it was fully phased in. A few years later, during a careful review of its labor-market projections, CBO (2014a) updated that estimate to roughly one percent, with the revision arising because the agency "incorporated into its analysis additional channels through which the ACA will affect labor supply, reviewed new research about those effects, and revised upward its estimates of the responsiveness of labor supply to changes in tax rates" (p. 118).

and making other changes to those tax systems. JCT published a set of reports analyzing that proposal, including estimates of the proposal's distributional consequences, revenue effects (excluding any impact on the overall economy), and overall economic impact. Based on the convention of fixed output, JCT (2014b) estimated that the proposal would be effectively revenue-neutral, raising federal revenues by $3 billion over the next decade. In its macroeconomic analysis, JCT (2014c, p. 21) wrote:

> Broadening of the individual and corporate income tax bases through elimination of many preferences in the form of deductions, exemptions, and tax credits allows for a reduction in average and effective marginal tax rates for most individual taxpayers, which provides both an incentive for increased labor effort, and an increase in demand for goods and services. These changes also reduce the after-tax return to investment under many modeling assumptions, providing an incentive for a reduction in the U.S. domestic capital stock. On net, these changes are expected to result in an increase in economic output relative to present law.

JCT estimated that the proposal would raise the level of output by between 0.1 percent and 1.6 percent, on average, during the 2014–23 period. That additional output was estimated to reduce cumulative deficits by between $50 billion and $700 billion during the 2014–23 period.

II.E. Other Illustrative Analyses

In the past several years, CBO and JCT have provided estimates of the budgetary feedback from the macroeconomic effects of other proposals as well.

Each year, the agencies publish detailed estimates of the president's budget proposals based on the conventional assumption that the overall economy would be unaffected, and CBO separately (and somewhat later) publishes an analysis of the economic effects of the proposals and the feedback to the federal budget. As an example, CBO (2012b, 2012c) estimated that, excluding macroeconomic effects, the cumulative deficits under the president's proposals would be $3.2 trillion during the 2013–17 period and another $3.2 trillion during the 2018–20 period—and that including macroeconomic effects, the cumulative deficits would be $3.0 trillion to $3.2 trillion during the first half-decade and $3.3 trillion to $3.6 trillion during the second half-decade.

In addition, CBO's annual analysis of the long-term budget outlook includes estimates of economic and budgetary outcomes under alternative policies, with the budgetary effects taking into account the economic effects and vice versa (CBO 2015e). Moreover, in some years, CBO has published

estimates of the economic effects and budgetary feedback of deficit paths specified by the chair of the House or Senate Budget Committee (for example, CBO 2015e). JCT has also released macroeconomic analyses of the Jobs and Growth Reconciliation Tax Act of 2003 (JCT 2003a), the American Recovery and Reinvestment Tax Act of 2009 as reported by the Ways and Means Committee (JCT 2009, pp. 234–39), and other proposals.

III. The Case for Including Macroeconomic Effects in Budget Estimates

A natural starting point for evaluating dynamic scoring is to consider the objectives of the budget estimating process. After providing that context, I explain why including macroeconomic effects in estimates for certain legislative proposals would both improve the accuracy of budget estimates for those proposals and provide important information about their economic effects. However, attempting to include macroeconomic effects in all budget estimates would not be appropriate because of limited staffing and time, and because it would raise another set of issues that would have to be addressed as well.

III.A. What Should Be the Objectives of the Budget Estimating Process?

In my judgment, *CBO and JCT should provide estimates for legislative proposals that measure the full budgetary effects as accurately as possible and illuminate notable nonbudgetary effects, subject to several significant constraints.*

To "measure the full budgetary effects as accurately as possible," estimates would ideally equal the expected present value of proposals' budgetary effects over a long time horizon. A long horizon is appropriate, because the effects of proposals could last into the indefinite future, and CBO and JCT should analyze proposals as specified and not presume the enactment of any future legislation to modify or undo them. In particular, it is untenable for agencies working for Congress to make specific predictions about the future decisions of current members or of members who will be elected over time.[10] Present value is appropriate because future events are

10. However, some approaches to estimating the macroeconomic effects of legislation do require limited predictions about future policies; this issue is addressed in section IV.

discounted in other contexts. The expected outcome is appropriate because it minimizes the mean squared error of estimates.[11]

To "illuminate notable nonbudgetary effects"—such as effects on the overall economy—estimates would ideally provide quantitative, or at least qualitative, information about such effects. That objective is appropriate because reliable and timely information about the nonbudgetary effects of proposals is important for policymakers to receive and is not readily available from sources other than CBO and JCT. In particular, advocates and opponents of proposals often generate overly optimistic or pessimistic estimates of their effects, while independent analysts often are not familiar with the details of proposals and do not possess the models needed to estimate their effects, so they have difficulty producing reliable estimates quickly.

However, those ideal approaches cannot be fully put into practice, due to four significant constraints, as follows:

First, estimates should be easily understandable by members of Congress, their staffs, and outside observers, and they should be resistant to misinterpretation. Many members, staff, and observers have little training in quantitative analysis or budgeting, and most have limited time for reviewing budget estimates. Also, advocates and opponents of proposals often try to cite estimates in ways that support their own positions, so it is important that estimates be clear and difficult to use in misleading ways.

Second, estimates should be based on methodologies that are applied consistently across related proposals and are credible to members of Congress, their staffs, and outside analysts. Using consistent methodologies is crucial to ensuring that proposals can be compared meaningfully. Using methodologies that are credible is crucial to maintaining Congress's confidence in the estimates, to ensuring that the estimates reflect the consensus of informed professional thinking, and to protecting CBO and JCT from political pressure. By contrast, using methodologies that seem arbitrary or can be easily manipulated by lawmakers' construction of proposals in particular ways undermines confidence in the agencies' estimates for those proposals and for other proposals as well.

Third, estimates should be produced quickly enough to serve the legislative process and structured in ways that fit the process. Thus, estimates should include the information sought by congressional leaders or committees as

11. For certain financial activities of the government where risk is apparent and can be readily assessed, I think the estimated budgetary effect should not equal the expected present value of the activities but instead should incorporate an adjustment for the cost of the risk. That issue lies beyond the scope of this paper.

they develop proposals and by members of Congress as they vote on proposals. And when the legislative process moves swiftly, estimates should be prepared and updated rapidly as proposals are modified.

Fourth, estimates need to be prepared using the resources available to CBO and JCT. Although the number of congressional requests for estimates has increased considerably in recent years, the funding provided to the agencies has left their staffing little changed, on balance.

Many aspects of the budget estimating process represent compromises between those constraints and the idealized estimates described above. As one important example, official budget estimates generally apply to the coming decade rather than a longer time period, because the estimating methodology needed for a longer period would require additional resources to develop, would usually be less credible, and would lead to estimates that were more prone to misinterpretation. However, certain proposals would have longer-term effects that are quite different from their effects in the coming decade, and in those cases CBO and JCT try to provide some information on longer-term effects. The precision of that information and the time period for which it is provided vary across proposals, depending on congressional interest, on the agencies' assessment of the resources required to generate the information, on the credibility of the methodology used, and on the risk of leaving results open to misinterpretation. Thus, CBO has analyzed certain proposals to change Social Security over 75 years (CBO 2010a), but it generally does not analyze proposals to change federal health care programs beyond 25 years because of the especially large uncertainty involved in predicting the evolution of the health care delivery and financing systems.

As another example, official budget estimates generally show nominal cash flows rather than inflation-adjusted or present-value cash flows because nominal flows are more straightforward than the alternatives and because the distortion relative to showing present values is fairly small over a decade. The principal exception is estimates for federal credit programs, for which nominal cash flows over a decade are often a gross misrepresentation of the full budgetary effects over a long horizon; for these estimates, accrued costs are therefore used instead.

As a final example, official budget estimates sometimes exclude factors that might affect the budgetary impact of proposals but whose sign or magnitude are especially uncertain. That exclusion may seem inconsistent with the objective of measuring budgetary impact as accurately as possible: The mean squared error of a budget estimate reflects the underlying uncertainty of all relevant factors, even if estimates of some of the factors are set to

zero; therefore, if CBO or JCT can generate informed estimates of those factors, including them in the overall budget estimate would probably increase the accuracy of the estimate. However, that point cannot be proven as a general rule. Although in-sample prediction errors from linear regression models are minimized by including all relevant factors, including additional factors does not necessarily minimize out-of-sample prediction errors from nonlinear models (which is what CBO and JCT often use).[12]

Moreover, including factors that are especially uncertain can diminish the credibility of the estimating process because when the likelihood function for a factor is particularly flat, the agencies' choice of a specific value often seems arbitrary. Therefore, for factors whose budgetary impacts are probably small and are especially uncertain, the probable improvement in accuracy from including them in budget estimates may be outweighed by the risks of inadvertently diminishing accuracy and weakening the agencies' credibility. That condition is particularly likely to be satisfied when the net budgetary impact of a set of excluded factors might be either positive or negative.[13]

III.B. Similarity between Macroeconomic and Nonmacroeconomic Effects

Changes in federal tax and spending policies can affect people's behavior in many ways, and those behavioral responses can affect the federal budget. Some of those responses affect the composition of output or distribution of income but not total output and income, while other responses affect total output and income as well as their composition and distribution. A natural presumption is that measuring the full budgetary effects of legislative proposals as accurately as possible requires including the impact of all of those behavioral responses.

For example, if marginal income tax rates were increased, a number of responses would ensue. The share of people's income devoted to activities whose costs can be deducted from income, such as mortgage interest

12. See CBO (2015f, pp. 21–22) for a related discussion.

13. For example, CBO (2014a, p. 123) discussed some ways in which the ACA might affect productivity and concluded: "Whether any of those changes would have a noticeable influence on overall economic productivity, however, is not clear. Moreover, those changes are difficult to quantify and they influence labor productivity in opposing directions. As a result, their effects are not incorporated into CBO's estimates of the effects of the ACA on the labor market."

payments and charitable contributions, would increase. The share of compensation received in nontaxable forms, such as employers' contributions to pensions and health insurance, would increase. The amount of labor supplied would decrease (if the substitution effect outweighed the income effect), and the amount of saving would decrease (again, if the substitution effect outweighed the income effect). The first two responses are typically included in conventional estimates, and the latter two are not. However, when the responses are described in this manner, there is no clear conceptual reason to treat the latter two responses differently from the first two.

Of course, one difference between those two sets of responses is that reductions in labor supply and saving would affect total output whereas shifts in the uses of income and types of compensation would not. Thus, the reductions in labor supply and saving can be labeled macroeconomic effects, while the other shifts are purely microeconomic. Still, because all of the responses stem from actions by people and firms, there is no clear rationale for including some in budget estimates and excluding others.

The reductions in labor supply and saving that are spurred directly by the increase in marginal tax rates could generate further economic changes, such as shifts in pretax wages and in the pretax return to capital, which would have further effects on labor supply and saving. One might argue that those additional effects should be excluded from budget estimates because of their indirectness. However, indirect effects can be quantitatively important. Consider an example from a legislative proposal that did not change tax rates: The immigration legislation approved by the Senate in 2013 would have significantly increased the supply of labor, which would have induced additional capital investment. Ignoring the increase in labor supply would have substantially understated the impact of the legislation on output; including that increase in labor supply without including the induced growth of the capital stock would still have understated the impact of the legislation on output and would also have overstated its impact on wages.

Changes in federal policies can affect total output and income in many other ways as well. Changes in tax rules can affect investment in human capital and the allocation of physical capital, changes in federal benefits can affect labor supply and saving, and changes in federal spending for infrastructure, education and training, and research and development can affect labor supply, saving, and productivity. As with the effects of changes in marginal tax rates, there is no clear conceptual basis for including in budget estimates the effects of such policy changes on specific parts of the economy but not the effects on aggregate economic variables.

III.C. Advantages of Including Macroeconomic Effects in Certain Budget Estimates

Based on CBO's and JCT's past analysis of the macroeconomic effects of legislative proposals and their budgetary feedback, I conclude that *using dynamic scoring in budget estimates for certain proposals would improve the accuracy of those estimates, provide important information about the economic effects of those proposals, and (under certain conditions) satisfy the significant practical constraints for budget estimates listed earlier.*

Some proposals' estimated macroeconomic effects would have significant budgetary consequences. For example, the estimated macroeconomic effects of the Senate's 2013 immigration bill, the ACA, and Congressman Camp's tax plan (based on the midpoint of the reported estimates) all have budgetary impacts equal to hundreds of billions of dollars over a decade. Moreover, if dynamic scoring had been applied to the economic stimulus legislation of 2009 (the American Recovery and Reinvestment Act), its estimated budgetary effect would have been reduced by hundreds of billions of dollars: CBO (2009b) estimated that the legislation would raise output by more than $800 billion over the following decade, and that additional income would have been estimated to reduce budget deficits by about $200 billion compared with an estimated budgetary cost of the bill of roughly $800 billion (CBO 2009a).[14]

To be sure, accounting for the estimated macroeconomic effects of those proposals would have improved the accuracy of the official budget estimates only if the estimates of the macroeconomic effects had been somewhat accurate.[15] Unfortunately, assessing the accuracy of CBO's and JCT's estimates is quite difficult. Many proposals that the agencies examined were not enacted, and the proposals that were enacted were just a few of many factors affecting the economy and the budget, so isolating their impact is hard even in retrospect (CBO 2013a, 2015b). In my judgment, however, both agencies' methodology for conducting macroeconomic analysis reflects the consensus of informed professional thinking, and that consensus provides a useful, albeit imperfect, basis for predicting the macroeconomic effects of legislative proposals. Moreover, there is no reason to believe that their

14. The central estimate in JCT (2009) was that the tax provisions in the bill (as approved by the Ways and Means Committee) would increase output by about one-half percent in the short run, leading to a reduction in the cost of those provisions of about one-seventh of the conventional estimate.

15. The further step of estimating the budgetary feedback from estimated macroeconomic effects is fairly straightforward and can be done reasonably accurately.

estimates of macroeconomic effects are generally less accurate than their estimates of other effects of complex proposals, although certainly the agencies should continue to strive to improve their analyses.

The effects of some legislative proposals on the overall economy are very important for policymakers to understand. For example, while the macroeconomic effects of immigration reform and tax reform are among the most touted reasons for pursuing those policy changes, different approaches to immigration reform—such as increasing the numbers of high-skilled and low-skilled immigrants by different amounts—could lead to very different macroeconomic effects.[16] Similarly, different approaches to tax reform, such as using revenues raised by broadening tax bases to reduce marginal tax rates or to make targeted inframarginal tax reductions, could lead to very different macroeconomic effects.

As another example, major changes to benefits for lower-income people could have notable effects on the economy by altering labor supply, and those effects could be an important criterion in evaluating such changes. To use Arthur Okun's famous metaphor, we should understand the leakiness of different buckets for transferring resources to lower-income people. And as a further example, policy changes that reduced federal deficits to different degrees and at different speeds would generally have different macroeconomic effects in the next few years and in the longer run.

Estimates of macroeconomic effects can be valuable even when those effects appear small to some observers. For example, CBO (2015e) found that this year's budget resolution—which calls for a reduction in cumulative deficits over the next decade of about $5 trillion excluding interest savings and macroeconomic effects—would raise the level of real output in 2025 by 1½ percent, which amounts to an increase in the average annual growth rate over the coming decade of 0.15 percentage point. If that effect is surprisingly small to some people, the value of the estimate is increased, not diminished.

In addition, objective and timely information about the macroeconomic effects of legislative proposals is not readily available from sources other than CBO and JCT. Advocates and opponents of particular policies usually find ways to have their perspectives well represented in the congressional and public debates. However, independent, reliable analysts generally

16. Changes in overall output do not necessarily correspond to changes in economic well-being and should not be interpreted as such. For example, CBO (2013c) distinguished carefully between the effects of the-Senate's 2013 immigration legislation on total output and on output per resident.

have more difficulty than the agencies do in completing their analyses on a timely basis (because doing careful analysis is hard and because those analysts often are not close to the policy development process) and having their analyses heard. Including macroeconomic effects in budget estimates would ensure that CBO and JCT devote resources to analyzing those effects in a timely way. Also, because of the attention paid to official budget estimates in the legislative process, including macroeconomic effects in those estimates would ensure that the effects received attention.

The usefulness of such attention depends in part on the clarity of CBO's and JCT's descriptions of estimated macroeconomic effects. Describing such effects can be challenging, but the agencies now have considerable practice doing so in their supplementary analyses and are quite able to do so in official budget estimates as well. For example, in the agencies' report on repealing the ACA, the logic and magnitude of the macroeconomic effects are laid out clearly and in a manner that nicely parallels the discussion of the repeal's nonmacroeconomic effects.

Lastly, under certain conditions CBO's and JCT's estimates of the macroeconomic effects of legislative proposals can satisfy the key constraints described earlier: being understandable and resistant to misinterpretation, based on a consistent and credible methodology, produced quickly enough to serve the legislative process, and prepared using the resources available to the agencies. Those issues are addressed in the remainder of the paper.

III.D. Limiting Macroeconomic Effects to Budget Estimates for Major Proposals

Despite the advantages of including macroeconomic effects in budget estimates for legislative proposals, I conclude that *such effects should be incorporated only in estimates for major proposals.* Specifically, apart from proposals for which dynamic scoring is requested by the chair or ranking member of the House or Senate Budget Committee, I think that dynamic scoring should be applied only to proposals whose estimated nonmacroeconomic effects on revenues, spending, or deficits, relative to the baseline, exceed a given threshold.

CONSIDERATIONS IN SUPPORT OF LIMITED DYNAMIC SCORING That recommendation is based on three considerations (although there are legitimate counterarguments that are discussed below). First, CBO and JCT have the resources to conduct careful macroeconomic analyses for only a limited number of legislative proposals each year. All of the estimates of macroeconomic effects described above involved significant conceptual and practical challenges and required a great deal of analysts' time to complete.

The agencies can devote that much time to only a very small share of the thousands of proposals they examine each year.

Second, applying rules of thumb to produce estimates of the macroeconomic effects of other proposals would generally violate the important constraint that estimates be based on methodologies that are credible and cannot be easily manipulated. The macroeconomic effects of proposals can be complex and can vary considerably with the specifics of the proposed policy changes, the state of the economy, and the time horizon being examined. For example, when CBO (2015c) examined three ways of reducing spending for the Supplemental Nutrition Assistance Program (commonly known as "food stamps"), the agency found that even the sign of some policies' net effect on labor supply was unclear without detailed analysis. Similarly, when JCT (2005) examined three approaches to reducing taxes by $500 billion, it found that their effects on the economy differed greatly.

In addition, the agencies' estimates of short-term macroeconomic effects depend importantly on the posture of monetary policy. Rules of thumb would not capture those kinds of crucial nuances. Moreover, developers of proposals might exploit rules of thumb by structuring their proposals or labeling aspects of their proposals in ways that would generate more favorable estimated macroeconomic effects and thus lower estimated budgetary costs. CBO and JCT currently minimize such gaming by basing their nonmacroeconomic estimates on a careful understanding of the substance of proposals rather than the application of arbitrary rules. Even if using rules of thumb to estimate macroeconomic effects could improve the accuracy of budget estimates on average, doing so would endanger the credibility of the estimating process.

Third, the proposals for which CBO's and JCT's estimates of macroeconomic effects would generally be most valuable are the ones with the largest estimated budgetary impacts apart from such effects—because those proposals are likely to produce significant macroeconomic effects. To be sure, *some* proposals that would not cause large changes in revenues or spending would also produce significant macroeconomic effects, but identifying them with a mechanical algorithm would be difficult. For example, Congressman Camp's comprehensive tax plan had a very small estimated effect on revenues apart from macroeconomic effects—because the large estimated effects of some individual provisions of the plan were largely offsetting—but significant estimated macroeconomic effects. A threshold for dynamic scoring based on the *gross* budgetary effects of a proposal's individual provisions would have identified his plan. However, a criterion based on so-called gross effects would not be very robust, because the

method of constructing an estimate and the level of detail shown in a table describing the estimate can alter the magnitude of the reported increases and decreases.[17]

Thus, the best way to choose which proposals with small estimated budgetary impacts relative to the baseline should be scored dynamically is to allow for requests from key congressional leaders. In principle, at least, the budget committees are responsible for the budget process, and the official budget estimates are designed to support that process, so it makes sense to allow for requests from those committees. One might also allow for requests from the House Ways and Means Committee, the Senate Finance Committee, and the House and Senate Appropriations Committees, which are the other principal consumers of CBO's and JCT's estimates. However, if all of those committees could request dynamic estimates, the number of requests might become unmanageably large, so I slightly prefer to vest the authority only in the budget committees. What is more important is to ensure that such requests are not skewed in the direction of any one party's political interests. Therefore, the ranking members as well as the chairs of the committees should be allowed to make requests. By contrast, the new House rule and this year's budget resolution grant that power only to the budget committee chairs.

In my view, a sensible threshold for automatically including macroeconomic effects in budget estimates would be estimated changes in revenues, spending, or deficits (excluding any macroeconomic impact) relative to the baseline exceeding one-quarter of one percent of projected output over the 10-year budget window. That threshold equals about $575 billion currently (based on CBO [2015a]) and probably would lead to dynamic scoring for only a few proposals each year, which would be a manageable increase in CBO's and JCT's workloads. Neither the Senate's 2013 immigration proposal nor Congressman Camp's tax plan would have met that threshold, though clearly at least one leader of a budget committee would have requested dynamic scoring for each.

The new House rule and this year's budget resolution use a quarter-point threshold but apply it to any single year in the budget window rather than

17. For example, the ACA included significant changes to the drug benefit in Medicare, some of which increased federal spending and others of which reduced it. One might view the estimated net effect on spending of those changes to be the combination of a gross estimated increase and a gross estimated decrease. However, the changes interacted with each other in significant ways, so CBO estimated their effects as a package and never identified elements of the estimate separately.

the 10-year period as a whole. That approach generates thresholds of about $45 billion in 2015 and about $70 billion in 2025 based on projected output, and naturally it will cause more bills to be scored dynamically than my preferred approach would. In particular, short-term extensions of expiring tax or spending provisions are more likely to receive dynamic scores, as in JCT (2015b). However, the estimated macroeconomic effects of such extensions are not always illuminating: Many people expect such provisions to be extended, even retroactively, but the baseline reflects the expiration of the provisions and therefore the assumption that people will gradually recognize that extensions are not occurring; as a result, the macroeconomic effects of extending the provisions are primarily the effects of people not being surprised. Such effects can be complicated to estimate and difficult to explain. In addition, extensions of that sort are often negotiated and voted on under tight timetables, which further complicates doing macroeconomic analysis.

TWO COUNTERARGUMENTS There are two noteworthy counterarguments to limiting dynamic scoring to major proposals. One is that even if the macroeconomic effects of a proposal with limited budgetary impact are small relative to the overall economy, their feedback effect on the federal budget could still be large relative to the nonmacroeconomic budgetary impact of the proposal. In those circumstances, careful dynamic scoring would significantly improve the accuracy of the budget estimate. However, CBO and JCT cannot do careful analyses of the macroeconomic effects of all proposals, and, as described above, using rules of thumb in place of careful analyses could reduce the accuracy of those estimates and diminish the credibility of CBO's and JCT's estimates more generally. In my judgment, those costs outweigh the benefits.

The other counterargument is that focusing dynamic scoring on major proposals would create an incentive for certain proposals to be bundled together or separated into pieces in order to lower their estimated budgetary cost. However, sufficiently few proposals have budgetary impacts close to a quarter of a percent of output that this distortion would probably not be significant in practice. Moreover, allowing key congressional leaders to request dynamic scoring for less-significant proposals should ameliorate this problem.

Given the inability of CBO and JCT to apply dynamic scoring to all proposals, one might wonder whether avoiding dynamic scoring altogether is the best feasible approach because it would make the estimating methodology more consistent across proposals. However, policymakers do not usually compare major proposals to less-significant proposals; major proposals

are usually compared to other major proposals addressing similar issues, or else they are compared to the status quo, and both types of comparisons would be improved by dynamic scoring. In addition, as just noted, key congressional leaders could request dynamic scoring for less-significant proposals whose comparison to major proposals they considered especially valuable.

III.E. Other Important Issues

Five other issues concerning the inclusion of macroeconomic effects in budget estimates deserve comment.

TIME HORIZONS First, estimates of macroeconomic effects of proposals should include both short-term effects stemming from shifts in aggregate demand and longer-term effects stemming from shifts in potential output. Longer-term economic effects may be better guides to proposals' effects beyond the 10-year budget window, and since that window is a compromise between the ideal of an even longer horizon and the practical advantages of a shorter horizon, a focus on longer-term effects may seem preferable. However, the severe recession and slow recovery of the past several years are a stark reminder that shortfalls in the demand for goods and services can have large and persistent effects on the economy and the federal budget, so the effects of policy changes on aggregate demand can be quite important. Increasing aggregate demand was the principal objective of some proposals considered by Congress in the past several years, such as the economic stimulus legislation of 2009.

In addition, policy changes are sometimes reversed or modified in subsequent years, so the short-term effects of changes are the effects most likely to occur, and policymakers may therefore give estimates of those effects greater weight in their decisions. Further, different policy changes that Congress sometimes compares—such as different time paths for reducing budget deficits by a given amount—would have different effects on aggregate demand, and illuminating those differences would be an important benefit of dynamic scoring.

RESOURCE CONSTRAINTS Second, when CBO and JCT conclude that they do not have the tools or time needed to do a careful analysis of a proposal's macroeconomic effects, they should state as much and not include such effects in the official budget estimate. Estimating macroeconomic effects carefully often requires a great deal of analysis, and legislation is sometimes developed and amended quickly. Doing dynamic scoring without sufficient tools and time would endanger the credibility of the estimating process, as discussed above. That situation is especially likely to arise for

changes in regulatory policy that have small effects on the federal budget apart from any macroeconomic effects. For example, substantial changes in federal regulation of the financial system or the environment could have significant macroeconomic effects that would be important for policymakers to understand and that could feed back to the federal budget in notable ways. However, CBO does not have much expertise in estimating the macroeconomic effects of such regulatory changes, and acquiring enough expertise to do so quickly during the legislative process would require a significant increase in CBO's resources and would distract the agency from its core responsibility of informing budget policy.

SHARED RESPONSIBILITY Third, CBO and JCT should share responsibility for estimating the macroeconomic effects of proposals along the lines of their sharing responsibility for estimating the budgetary impact of proposals excluding macroeconomic effects. Specifically, JCT should produce estimates of the macroeconomic effects of major proposals to change the federal tax code, CBO should do the same for major proposals to change federal spending, and the two agencies should collaborate in estimating the macroeconomic effects of major proposals that would change both tax and spending policies (as they did in their estimate for repealing the ACA, released in June 2015 [CBO 2015h]). One challenge is ensuring that proposals that are similar in their substance receive similar estimates of their macroeconomic effects regardless of whether they are structured as changes in tax policy or spending policy. For example, similar subsidies for similar activities should be estimated to have similar macroeconomic effects, whether those subsidies take the form of tax credits or explicit federal spending. But that same challenge arises currently in estimating the budgetary impact of policies excluding their macroeconomic effects, and the challenge is met by ongoing interaction and coordination between CBO and JCT.

CONGRESSIONAL PROCEDURES Fourth, some congressional procedures may need to be modified to accommodate dynamic scoring. For example, dynamic scoring may generate jurisdictional conflicts between some congressional committees. Because changes in spending that affected total output and income would affect revenues, the House and the Senate would need to develop procedures for assigning budget targets to committees that allowed for such interactions. That problem would be somewhat ameliorated by the limited number of bills that would be scored dynamically. Other congressional procedures would adapt naturally to dynamic scoring. For example, official budget estimates are used to assess the applicability of certain parliamentary "points of order"; if a proposal were scored dynamically, the

estimated budgetary feedback from the macroeconomic effects of the proposal would be incorporated in that assessment.

SUPPLEMENTARY REPORTS Fifth, there are some advantages to taking an alternative approach, in which CBO's and JCT's estimates of macroeconomic effects and their budgetary feedback would be provided in supplementary reports rather than being included in official budget estimates, although I think those advantages are outweighed by the approach's disadvantages. The advantages of that alternative approach include the following: It would avoid delaying the publication of budget estimates until macroeconomic analyses could be completed, which might be a considerable period of time in some cases. It would maintain a consistent basis of fixed output for all official budget estimates, which might make it easier for CBO and JCT to exclude macroeconomic effects when they did not have a solid analytic basis for assessing them. And it would avoid the jurisdictional problems that can arise between committees just discussed.

One might view as a further advantage of the alternative approach that estimated macroeconomic effects and their budgetary feedback could be reported as ranges of possible outcomes rather than as the point estimates that the congressional budget process requires. This would also demonstrate the uncertainty of such macroeconomic analyses. Indeed, CBO and JCT should quantify that uncertainty by reporting ranges of estimates whenever feasible, as I discuss in greater detail later. However, CBO (2014f, p. 12) explains that "providing ranges sometimes muddies, rather than enhances, general understanding of our analysis because people tend to cite the part of a range they prefer," so CBO already tries to clarify the agency's findings by reporting point estimates as well as ranges in the agency's macroeconomic analyses. Using those point estimates in official budget estimates would not distort the analyses or their presentation.

One might view as a different advantage of the alternative approach that it would enable the House and Senate to disagree about whether to include macroeconomic effects in official estimates and would enable members of Congress, their staffs, and outside observers to evaluate the estimated macroeconomic effects separately from the other estimated effects. However, this would already be the case without the alternative approach. CBO (2015d, p. 23) has explained that "cost estimates [with dynamic scoring] will include all of the information that typically would be included if macroeconomic effects were not incorporated in the analysis, as well as additional information related to the macroeconomic effects." Accordingly, the recent estimates for repealing the ACA and extending certain expiring tax provisions reported three projections: the estimated budgetary impacts

excluding macroeconomic effects, the estimated budgetary impacts of macroeconomic effects alone, and the estimated total budgetary impacts including macroeconomic effects. Including estimated macroeconomic effects in official budget estimates in this manner does not preclude Congress and others from evaluating and using that information as they see fit.

Yet another seeming advantage of the alternative approach might be to facilitate a procedural transition between the historical exclusion of macroeconomic effects and their possible future inclusion. Such a transition would give CBO and JCT an opportunity to experiment with different methods of analysis and presentation, and it would give Congress and others an opportunity to learn about the agencies' macroeconomic analysis and develop procedures for using that information. In fact, however, that transition has effectively been under way for some time. As discussed above, both CBO and JCT have published many analyses of the macroeconomic effects of legislative proposals as well as reports on their methodology for such analyses. Moreover, the recently published estimates for repealing the ACA and extending expiring tax provisions did not reveal any problems that would suggest dynamic scoring is "not ready for prime time."

In addition, the alternative approach would have some significant disadvantages. Leaving macroeconomic effects out of budget estimates would reduce the attention those effects receive, even though the effects can be important for policymakers to understand. Also, separating the two types of budgetary effects of legislative proposals would make CBO's and JCT's estimates more difficult to understand for members of Congress, their staffs, and outside observers.

In my judgment, all those disadvantages of the alternative approach outweigh its limited advantages.

IV. Concerns about Including Macroeconomic Effects in Budget Estimates

A number of observers have expressed the view that including the macroeconomic effects of legislative proposals in official budget estimates would worsen rather than improve the information those estimates provide. In this section I examine six important concerns noted by those observers that were not addressed, or were addressed only briefly, in the preceding section. In my view, the first two of these concerns apply with roughly equal force to macroeconomic and nonmacroeconomic effects of proposals and thus are not compelling reasons for treating the former differently,

the next three concerns represent true challenges in estimating macroeconomic effects but can be addressed adequately by CBO and JCT, and the final concern is premised on an incorrect view of the agencies' role in the budget process.

IV.A. Potential Political Pressure on CBO and JCT

One concern is that including macroeconomic effects of proposals in official budget estimates would increase political pressure on CBO and JCT to adopt modeling approaches or elasticities of behavioral responses that would overstate the positive effects of certain types of policies. I welcome vigilance to guard against the risk that the agencies will be subjected to political pressure to modify their analysis in any regard. However, political pressure has not altered CBO's or JCT's estimates for legislation in past decades, and I doubt that dynamic scoring would substantially increase such pressure.

In my six years as CBO director, members of Congress, their staffs, and other observers were not hesitant to speak up when they disagreed with an estimate from CBO. That is not surprising, nor is it objectionable, because CBO's and JCT's estimates are not above reproach. The appropriate response by the agencies to such criticism is to collect any information that can be provided by those who disagree with the estimate, reconsider whether the agencies' analysis was correct, and revise the estimate if, and only if, the reconsideration shows that a different figure would be more accurate. That process occurred a number of times while I was at CBO. However, at no point in those six years did anyone in a position of authority in Congress attempt to dictate a change in a CBO estimate or in a CBO modeling assumption. Congressional leaders appear to understand the long-term value of maintaining CBO's and JCT's analytic independence.

Similar disagreements with CBO's and JCT's macroeconomic analyses have occurred in the past and would occur in the future if dynamic scoring were adopted. However, the risk of political pressure does not seem greater for estimates of macroeconomic effects than for estimates of nonmacroeconomic effects. On the one hand, there may be more observers who have strongly held views about key parameters underlying macroeconomic estimates—such as the elasticity of labor supply—than have strongly held views about the less well-known parameters underlying nonmacroeconomic estimates. On the other hand, because macroeconomic estimates depend heavily on a small number of parameters and other modeling choices, CBO and JCT have publicly documented those choices more thoroughly than they have publicly documented the analytic underpinnings of some

nonmacroeconomic estimates. It is important, then, that CBO and JCT report publicly on any changes in their macroeconomic modeling (as CBO did during the past few years when it updated a number of aspects of that modeling).

IV.B. Uncertainty of Macroeconomic Effects

Another concern is that the macroeconomic effects of proposals are so uncertain that including them in official budget estimates would degrade the quality of the estimates. However, as noted above, the accuracy of the budget estimates would probably be improved by including all of the factors for which CBO or JCT can generate informed estimates, and macroeconomic effects meet that criterion when the agencies have the time and tools to do a careful analysis. Moreover, many nonmacroeconomic effects of proposals are very uncertain as well, and in many cases there is less evidence to use in quantifying those effects than in quantifying the macroeconomic effects of proposals.

As one important example, the extent of uncertainty regarding the macroeconomic effects of changes in tax rates can be gleaned from CBO's recent review of evidence of the elasticity of labor supply. In a table summarizing estimates of the substitution elasticity for men and single women, CBO (2012g) showed values ranging from 0.04 to 0.84; in a corresponding table for married women, CBO showed values ranging from 0.03 to 0.70. Similar uncertainty exists about other behavioral responses and other aspects of CBO's and JCT's models of the macroeconomic effects of fiscal policies. However, uncertainty about those parameters is so apparent in part because they have been the subject of substantial research, which at least provides evidence for CBO and JCT to draw on.

For many of the behavioral responses underlying the nonmacroeconomic effects of fiscal policies, there is much less evidence for CBO and JCT to use, but that does not imply that uncertainty about the responses is smaller. For example, the agencies' estimates of the nonmacroeconomic effects of the Senate's 2013 immigration legislation, the ACA, and Congressman Camp's tax reform plan all relied heavily on assessments of behavioral responses for which there is little evidence and experts are highly uncertain. In addition, as noted above, the agencies' public documentation of their methodologies for estimating nonmacroeconomic effects is less expansive in some cases than their public documentation of their methodology for estimating macroeconomic effects. Thus, there is no good reason to view the agencies' estimates of macroeconomic effects as less credible than their estimates of nonmacroeconomic effects.

Note also that *excluding* macroeconomic effects from budget estimates for proposals that might have significant macroeconomic effects—and doing so only because of a historical convention that many consider arbitrary—itself diminishes the credibility of budget estimates. In sum, I think that including macroeconomic effects (under the conditions described in this paper) would enhance the credibility of the official budget estimates.

Although the budget process focuses on point estimates, quantifying the uncertainty of estimated macroeconomic effects may be useful to members of Congress, their staffs, and outside analysts, and I think that CBO and JCT should do that whenever feasible.[18] For example, the range of estimates that JCT published for Congressman Camp's tax plan provided a measure of uncertainty, and, as noted above, CBO typically presents its estimates of macroeconomic effects with a range as well as a central estimate.

IV.C. Potential Bias toward Tax Cuts Relative to Spending Increases

A further concern is that applying dynamic scoring to proposals affecting federal taxes but not proposals affecting federal spending would distort policymakers' decisions in favor of tax cuts relative to spending increases. For example, if lower tax rates raise output by increasing labor supply, and greater infrastructure spending raises output by increasing the capital stock, then including the former effect in official budget estimates but excluding the latter effect would inappropriately encourage tax rate cuts relative to infrastructure spending increases. That concern is, in some ways, the opposite of a concern about the conventional approach to cost estimates, namely that excluding effects on labor supply tends to overstate the budgetary cost of tax cuts and understate the budgetary cost of benefit increases, thereby encouraging policymakers to increase taxes and benefits. In any event, the concern can be addressed by applying dynamic scoring to proposals that change spending as well as those that change revenues.

Indeed, as described earlier, CBO's analyses of macroeconomic effects include the effects of federal spending on the demand for goods and services, the effects of federal benefits on labor supply, and the effects on the economy of federal investments in infrastructure, education and training,

18. Regarding uncertainty in CBO's estimates and the appropriate response by policymakers, see Manski (2011), CBO (2014f), and CBO (2015g, pp. 108–9).

and research and development.[19] For example, CBO's annual economic analysis of the President's budget incorporates the effects of proposed changes in both spending and taxes. In addition, the estimated effects of the ACA on labor supply stem partly from changes in the tax code and partly from changes in spending for Medicaid (as well as some other aspects of the law), and those effects are treated in a completely parallel manner in the estimates.

Nonetheless, there are two obstacles to the goal of applying dynamic scoring equally to federal spending changes and tax changes. One obstacle is that the congressional budget process treats certain types of spending differently from other types of spending and revenues. Roughly a third of noninterest federal spending arises from annual appropriations by Congress (sometimes called "discretionary spending"), with the remaining roughly two-thirds reflecting payments for ongoing benefit programs (sometimes called "mandatory spending"). Appropriations are currently split about equally between defense and nondefense purposes, and about half of nondefense appropriations go to investments in infrastructure, education and training, and research and development. The Congressional Budget Act of 1974—which established CBO, the House and Senate Budget Committees, and many of the ground rules that govern the budget process—specified that CBO should not produce estimates for appropriations bills that are comparable to those it produces for other legislation, but instead should tally the amounts specified in those bills and provide those tallies to the appropriations committees. Potential implications of appropriations for future tax revenues or benefit payments are not considered in that tallying process or in the subsequent legislative process. Perhaps because of that different procedural treatment, appropriations bills are excluded from the requirements for dynamic scoring in the new House rule and in the congressional budget resolution.

Moreover, if dynamic scoring were applied only to proposals with a significant budgetary impact (excluding macroeconomic effects) relative to the

19. As CBO has noted, its analysis of those effects would benefit from further methodological advances. CBO (2014d, p. 9) explained that the agency is "developing the capability to apply substitution elasticities as well [as income elasticities], but it does not currently have that capability for all transfer payments (although the agency has incorporated substitution elasticities in some specific analyses)." Similarly, the agency should enhance its capability to estimate the effects of federal investments in disaggregated categories and to estimate the long-term effects of federal benefits—although its analysis in those areas is limited mostly by a paucity of available research.

baseline, as recommended above, few appropriations bills would meet that criterion. The baseline for appropriations equals the previous year's appropriations adjusted for inflation, or the statutory cap if one exists. Actual appropriations in a single year rarely differ from the baseline by an amount that exceeds a quarter of a percent of output over the coming decade, with the most recent exception being the one-time burst of appropriations under the 2009 stimulus bill.

However, if the chairs and ranking members of the budget committees were allowed to request dynamic scoring for some bills with small estimated budgetary effects, also as recommended above, then they should be allowed to make those requests for appropriations bills. It would be useful, in my view, for CBO to provide estimates of the macroeconomic effects and resulting budgetary feedback of consequential changes in appropriations.[20] Unfortunately, the new House rule and congressional budget resolution explicitly exclude appropriations bills. I also think that CBO should publish a report with estimates of the macroeconomic effects of alternative multiyear paths for federal investment and the budgetary feedback from those macroeconomic effects.

A second obstacle to applying dynamic scoring equally to spending changes and tax changes is that the macroeconomic effects of certain sorts of federal spending are not fully felt within the 10-year budget window. Of course, that same issue arises for certain sorts of federal tax changes. For example, reductions in the marginal tax rate on capital income encourage additional private investment, and the resulting increase in the capital stock (and thus output) occurs gradually. However, the problem may be especially acute for spending changes that involve investments. For example, most of the increment to output from a new bridge or improved highway (allowing for construction time and subsequent depreciation) occurs beyond the budget window. And most of the increment to output and income that may arise from improved health care, preschool education, or housing for low-income children occurs after those children have entered the labor force. Therefore, the extent to which the budget window distorts the estimated macroeconomic effects of a change in spending (or tax) policy varies considerably depending on the characteristics of the policy change and the private or public investment that is increased or decreased.

This problem can be at least partly addressed by having CBO and JCT provide information about the effects of proposals beyond the 10-year budget

20. For a recent estimate of the macroeconomic effects of a change in appropriations—but not the budgetary feedback from those macroeconomic effects—see CBO (2015l).

window. As noted above, when Congress is especially interested in a proposal's long-term effects or when CBO or JCT expect that a proposal's long-term effects would be very different from its effects within the budget window, the agencies can provide that information. For example, CBO and JCT provided estimates for the second decade after enactment for the Senate's 2013 immigration legislation and for repeal of the ACA, and in some cases CBO has estimated the effects of policy changes over even longer horizons (CBO 2010a, 2012e). In addition, the new House rule and congressional budget resolution explicitly require that qualitative information be provided about budgetary effects (including macroeconomic effects) beyond the 10-year budget window. Unfortunately, providing such information beyond the coming decade is challenging because, as noted earlier, the estimating methodology needed for a longer period would require additional resources to develop, would usually be less credible, and would lead to estimates that were more prone to misinterpretation in certain ways. In particular, the evidence base that could be used to construct estimates of the long-term effects of benefit changes for low-income children remains limited.

Given the limitations on applying dynamic scoring to changes in federal spending, one might wonder whether avoiding dynamic scoring for changes in federal taxes is the best feasible approach to official budget estimates because it increases the comparability of certain policies in a second-best way. Indeed, one might wonder whether dynamic scoring for changes in taxes always provides useful information if some of the macroeconomic effects of those changes occur beyond the 10-year budget window. However, the macroeconomic effects of policy changes within the budget window can be important, even if later effects are somewhat different. As noted above, the severe recession and slow recovery of the past several years, and the fact that policy changes are sometimes reversed or modified over time, may make policymakers especially interested in short-term effects.[21]

In addition, the limitations discussed here do not create significant distortions in comparisons among many types of tax changes, comparisons between tax changes and the continuation of current tax law, comparisons between tax changes and spending changes that do not have notable effects on investment, or comparisons between spending changes apart from investment and the continuation of current law. Even for comparisons

21. The particular problems that might arise in providing estimated macroeconomic effects for unsustainable policy changes are discussed in a later section.

between tax changes and spending changes that affect investment, estimates for the decade-long budget window or estimates that extend into a second decade may still be more comparable if they include macroeconomic effects. At the least, including such effects in budget estimates, even when imperfectly measured, would bring additional attention to those effects.

IV.D. Difficulty in Capturing Expectations of Future Fiscal Policy

Yet another concern is that accurately projecting the macroeconomic effects of proposals might require modeling expectations of future fiscal policy, which would raise significant conceptual and practical difficulties. In particular, a reduction in current tax rates may have different effects on labor supply and saving if people's expectations of future tax rates increase, decrease, or remain the same—and modeling those expectations is challenging, especially if the proposal at hand represents an unsustainable change in policy. However, expectations of future fiscal policy are frequently left aside in other economic analyses, and CBO and JCT have developed methods for handling this issue in their macroeconomic analyses.

Consider proposals that reduce marginal tax rates on labor income. When CBO or JCT analyze such a proposal using their Solow-type growth models, they apply labor supply elasticities drawn from the large empirical literature to the change in tax rates between the baseline and the proposal. Many papers in that literature do not explicitly measure expected tax rates but simply examine changes in labor supply that have resulted from given changes in contemporaneous tax rates. Therefore, the estimated elasticities can be interpreted as the effect of a given change in current tax rates with expected future tax rates adjusting in whatever way people expected them to adjust, on average, in the past—which may appropriately reflect the fuzziness of people's expectations about future tax rates. The Solow model does not explicitly include expectations, so this sort of estimated elasticity fits logically, although it will generate a less accurate estimate in circumstances when anticipatory effects are important.

In contrast, people's behavior in life cycle growth models depends explicitly on their expectations. If forward-looking people expected that federal debt would rise relative to output without limit, they would not hold federal bonds, so the models can be used only to analyze sustainable changes in policies. Therefore, when CBO or JCT use their life cycle models to analyze a proposal that would increase deficits indefinitely, the agencies incorporate future policy changes not specified in the proposal to offset the deficit increases. That situation is awkward, because a key principle of budget

estimates is that the agencies take proposals as written and do not predict future legislation.

To minimize the influence of the assumed future policy changes on their estimates, CBO and JCT generally report results for multiple alternative changes—for example, one estimate under the assumption that future revenues are increased and another under the assumption that future spending is reduced. In fact, the results often do not differ very much under alternative assumptions (CBO 2015i). The agencies also defer the assumed changes as long as possible while still being able to solve the models (although congressional interest in receiving information about the effects of proposals beyond the 10-year budget window increases the difficulty of deferring the assumed changes long enough that they would not affect the reported estimates). Even so, if a proposal would increase deficits indefinitely, which would have harmful economic effects, the inclusion in a budget estimate of additional policy changes not specified in the proposal might make the proposal look better than it really is.

Given those issues, I think that CBO and JCT should give less weight to estimates based on their life cycle models than to estimates using their Solow-type models, except in circumstances where the anticipatory effects of proposals might be especially important.[22] Indeed, the agencies have used their life cycle models less often than their Solow-type models in their macroeconomic analyses, perhaps because of the issue described here or else because the life cycle models can be more cumbersome in other respects.

It bears emphasis that the dependence of behavior on expectations of future fiscal policy also arises with estimates of certain nonmacroeconomic effects of proposals, although it is rarely discussed in that context. For example, a reduction in current tax rates may have different effects on mortgage interest deductions or on employers' payments for health insurance, depending on expectations of future tax rates. However, the empirical literature regarding such behavior does not explicitly measure expected tax rates, so CBO's and JCT's estimates can be interpreted as the effects of a given change in current tax rates with expected future tax rates adjusting in whatever way people had expected them to adjust, on average, in the past. That approach is somewhat unsatisfying, especially if the policy change at hand is not sustainable and therefore future changes will be needed, but there is no feasible alternative.

22. See CBO (2014c, pp. 14–15) for an example of how the agency combines results from its two models.

IV.E. Potential Distortion to Estimates of Unsustainable Policy Changes Stemming From the 10-Year Budget Window

A further concern is that estimates of the macroeconomic effects of unsustainable policy changes would be distorted by the 10-year budget window. For example, suppose that a reduction in tax rates was estimated to raise output over the next decade but also to generate revenue losses that were unsustainable (because the increase in output was not large enough for the tax reduction to pay for itself). If those same rates were later raised to satisfy the government budget constraint, output would be estimated to be lower in the long run. In that scenario, dynamic scoring over the budget window would credit the proposal with raising output even though the ultimate effect on output would be negative.

That scenario might occur, but it is not likely. First, a reduction in tax rates without an offsetting broadening of the tax base or reduction in spending might well be estimated to *lower* output within the 10-year budget window, depending on the specifics of the tax reduction. For example, JCT (2003a) estimated that the tax cuts enacted in 2003 would increase output during the first five years after enactment but decrease output later in the decade, in part because the harmful effects of greater federal debt were estimated to outweigh the favorable effects of lower tax rates. Similarly, CBO (2010c) estimated that extending the tax cuts originally enacted in 2001 and 2003 would raise output in the following few years but lower output later in the decade.[23]

Second, the harmful effects of greater federal debt increase over time as debt compounds, while the favorable effects of lower tax rates generally do not, so any tax-rate reductions that were estimated to raise output throughout the first decade after enactment would be less likely to be estimated to do so in the second decade. Indeed, CBO (2010c) estimated that the negative effects on output of extending the tax cuts would be much larger after 30 years than after 10 years. Those long-term effects would be reported by CBO and JCT because, as noted above, the agencies try to provide information about long-term effects when they expect them to be very different from effects within the budget window. In addition, the

23. That result may seem surprising in light of CBO's (2005) estimate that a 10-percent reduction in federal tax rates on individual income would probably increase output in the second half of the decade after enactment. The difference between CBO's 2005 and 2010 estimates stems from several factors, including: the inclusion of inframarginal cuts in taxes under the 2001 and 2003 legislation; an increase in outstanding federal debt, which means that the rise in interest rates resulting from greater federal borrowing has a more significant effect on future deficits; and various improvements in CBO's modeling.

current rules regarding dynamic scoring require CBO and JCT to provide information about the long-term effects of proposals.

Despite those points, suppose that a reduction in tax rates was, in fact, estimated to raise output over the next few decades but also to generate revenue losses that were unsustainable. The unsustainable nature of the rate reduction would be shown by the estimated effect of the proposal on federal debt, so the need to make further policy changes to offset the budgetary losses would be quite apparent. In addition, the increase in deficits might be offset later *not* by reversing the tax-rate reduction but by making some other policy change—and because that other change would have an effect on output that was not simply the opposite of the effect of the tax-rate reduction, it is unclear whether the estimated effects of the rate reduction on output in the first few decades would truly be misleading. Moreover, the estimated macroeconomic effects of the rate reduction would not be the only aspect of the budget estimate that could be misleading; the estimated nonmacroeconomic effects on the budget could be misleading as well, as would any distributional analysis or other analysis based on the rate reduction. In any event, it is not tenable for CBO and JCT to ignore the policy changes included in a legislative proposal even if those changes are not, by themselves, sustainable.

IV.F. Potential Benefits of Estimates that Err on the Side of Overstating Budgetary Costs

One other concern about dynamic scoring arises from the view that policymakers tend to give insufficient weight to budgetary costs when developing and voting on legislative proposals, so CBO and JCT should provide budget estimates that tend to err in the direction of overstating those costs. As summarized (but not necessarily endorsed) by the Committee for a Responsible Federal Budget (2012, p. 1), the argument is then made that excluding the macroeconomic effects of proposals from official budget estimates would make "dynamic gains a 'bonus' to help further reduce the deficit."

However, dynamic scoring does not consistently reduce the estimated budgetary cost of proposals relative to nondynamic scoring. As noted earlier, CBO (2010c) estimated that extending the broad tax cuts originally enacted in 2001 and 2003 would *reduce* output by the latter part of the decade after enactment, so dynamic scoring would have shown a more negative impact on the budget than nondynamic scoring. Also, CBO (2014a) estimated that the expansion of federal subsidies for health insurance under the ACA was reducing labor supply and thereby federal revenues, so dynamic scoring of

the ACA would have shown a more negative impact on the budget than nondynamic scoring.

More fundamentally, I do not think it is appropriate for CBO and JCT to try to nudge policymakers toward smaller budget deficits by providing estimates that tend to overstate the budgetary costs of proposals. Instead, the agencies should provide estimates that are in the middle of the distribution of possible outcomes and leave policymakers to make decisions based on their own views of desirable outcomes and acceptable risks.

V. Conclusion

Including macroeconomic effects in budget estimates for major legislative proposals—except when CBO and JCT do not have the tools or time to do a careful macroeconomic analysis—would improve the accuracy of those budget estimates and would provide important information about the economic effects of those proposals. Therefore, I conclude that dynamic scoring should be used for major proposals when the agencies have the tools and time to do a careful macroeconomic analysis.

To complement that greater commitment to analyzing the macroeconomic effects of legislative proposals, CBO and JCT should also make a greater commitment to analyzing proposals' distributional effects. JCT currently provides estimates of the distributional consequences of certain changes in federal taxes, as it did when it analyzed Congressman Camp's comprehensive tax reform proposal (JCT 2014a). However, CBO does not provide corresponding estimates for changes in federal spending. CBO's recent reports on the distribution of federal taxes under current law (CBO 2014e) and the distribution of federal taxes and spending under current law (CBO 2013d) described a number of conceptual complications and data limitations that arise in estimating the distributional impact of existing taxes and spending, and those problems are more acute when estimating the distributional impact of *changes* in taxes and spending. For example, distributional analysis of proposals ideally would incorporate macroeconomic analysis to capture the partial shift in the burden of capital taxes from capital owners to workers stemming from changes in the amount of capital. However, CBO and JCT have not developed the models needed to conduct such "dynamic distributional analysis."

In my view, CBO and JCT should continue to enhance their capabilities in this area. Policy choices can have significant effects on the distribution of income as well as on total income, and outcomes that are quantified often receive greater attention in policy discussions than outcomes that are

not. Therefore, it would be very valuable for CBO and JCT to quantify the effects of legislative proposals on both total income and its distribution.

ACKNOWLEDGMENTS I am grateful to Thomas Barthold, Wendy Edelberg, Glenn Hubbard, Jeffrey Kling, Donald Marron, Pam Moomau, Ben Page, and the editors for helpful comments on earlier drafts of this paper and to many former colleagues at the Congressional Budget Office and on the staff of the Joint Committee on Taxation for discussions that helped to form my views on this topic. The views expressed here are my own and should not be attributed to any of those people or to any institution with which I was or am currently affiliated.

References

Auerbach, Alan J. 1996. "Dynamic Revenue Estimation." *Journal of Economic Perspectives* 10, no. 1: 141–57.

———. 2005. "Dynamic Scoring: An Introduction to the Issues." *American Economic Review* 95, no. 2: 421–25.

Burman, Leonard E. 2006. "Dynamic Analysis and Scoring." Statement before the House Committee on the Budget, September 13. Washington. http://www.taxpolicycenter.org/UploadedPDF/900999_dynamicanalysis.pdf

CBO (Congressional Budget Office). 1995. "Budget Estimates: Current Practices and Alternative Approaches." Washington. https://www.cbo.gov/publication/13599

———. 1999. "Estimating the Costs of One-Sided Bets: How CBO Analyzes Proposals with Asymmetric Uncertainties." Washington. https://www.cbo.gov/publication/11843

———. 2001. "CBO's Method for Estimating Potential Output: An Update." Washington. https://www.cbo.gov/publication/13250

———. 2002. "Federal Budget Estimating." Statement of Dan L. Crippen before the Subcommittee on Legislative and Budget Process, House Rules Committee, May 9. Washington. https://www.cbo.gov/publication/13648

———. 2005. "Analyzing the Economic and Budgetary Effects of a 10 Percent Cut in Income Tax Rates." Washington. https://www.cbo.gov/publication/17507

———. 2009a. "H.R. 1: American Recovery and Reinvestment Act of 2009." Cost Estimate. Washington. https://www.cbo.gov/publication/41759

———. 2009b. "Estimated Macroeconomic Impacts of H.R. 1 as Passed by the House and by the Senate." Letter to the Honorable Judd Gregg, Ranking Member, Committee on the Budget, United States Senate, February 11. Washington. https://www.cbo.gov/publication/20474

———. 2009c. "Effects of Changes to the Health Insurance System on Labor Markets." Washington. https://www.cbo.gov/publication/20910

———. 2010a. "Social Security Policy Options." Washington. https://www.cbo.gov/publication/21547

———. 2010b. "The Budget and Economic Outlook: An Update." Washington. https://www.cbo.gov/publication/21670

———. 2010c. "The Economic Outlook and Fiscal Policy Choices." Statement of Douglas W. Elmendorf before the Committee on the Budget, United States Senate, September 28. Washington. https://www.cbo.gov/publication/21836

———. 2011. "CBO's Use of Evidence in Analysis of Budget and Economic Policies." PowerPoint Presentation at the Annual Fall Research Conference of the Association for Public Policy Analysis & Management, November 3. https://www.cbo.gov/publication/42722

———. 2012a. "Fair-Value Accounting for Federal Credit Programs." Washington. https://www.cbo.gov/publication/43027

———. 2012b. "An Analysis of the President's 2013 Budget." Washington. https://www.cbo.gov/publication/43083

———. 2012c. "The Economic Impact of the President's 2013 Budget." Washington. https://www.cbo.gov/publication/42972

———. 2012d. "Assessing the Short-Term Effects on Output of Changes in Federal Fiscal Policies." Working Paper no. 2012-08. Washington. https://www.cbo.gov/publication/43278

———. 2012e. "Raising the Excise Tax on Cigarettes: Effects on Health and the Federal Budget." Washington. https://www.cbo.gov/publication/43319

———. 2012f. "How the Supply of Labor Responds to Changes in Fiscal Policy." Washington. https://www.cbo.gov/publication/43674

———. 2012g. "A Review of Recent Research on Labor Supply Elasticities." Working Paper no. 2012-12. Washington. https://www.cbo.gov/publication/43675

———. 2012h. "Review of Estimates of the Frisch Elasticity of Labor Supply." Working Paper no. 2012-13. Washington. https://www.cbo.gov/publication/43676

———. 2013a. "The Accuracy of CBO's Budget Projections." CBO blog post by Doug Elmendorf, March 25. Washington. https://www.cbo.gov/publication/44017

———. 2013b. "S. 744: Border Security, Economic Opportunity, and Immigration Modernization Act." Cost Estimate. Washington. https://www.cbo.gov/publication/44225

———. 2013c. "The Economic Impact of S. 744, the Border Security, Economic Opportunity, and Immigration Modernization Act." Washington. https://www.cbo.gov/publication/44346

———. 2013d. "The Distribution of Federal Spending and Taxes in 2006." Washington. https://www.cbo.gov/publication/44698

———. 2013e. "Fiscal Policy Effects in a Heterogeneous-Agent Overlapping-Generations Economy with an Aging Population." Working Paper no. 2013-07. Washington. https://www.cbo.gov/publication/44941

———. 2014a. "The Budget and Economic Outlook: 2014 to 2024." Washington. https://www.cbo.gov/publication/45010

———. 2014b. "The Long-Run Effects of Federal Budget Deficits on National Saving and Private Domestic Investment." Working Paper no. 2014-02. Washington. https://www.cbo.gov/publication/45140

———. 2014c. "The Economic Effects of the President's 2015 Budget." Washington. https://www.cbo.gov/publication/45540

———. 2014d. "How CBO Analyzes the Effects of Changes in Federal Fiscal Policies on the Economy." Washington. https://www.cbo.gov/publication/49494

———. 2014e. "The Distribution of Household Income and Federal Taxes, 2011." Washington. https://www.cbo.gov/publication/49440

———. 2014f. "Communicating Uncertainty in Budgetary and Economic Estimates." Presentation by Douglas W. Elmendorf at the Hutchins Center on Fiscal

and Monetary Policy at the Brookings Institution, December 15. Washington. https://www.cbo.gov/publication/49862

———. 2015a. "The Budget and Economic Outlook: 2015 to 2025." Washington. https://www.cbo.gov/publication/49892

———. 2015b. "CBO's Economic Forecasting Record: 2015 Update." Washington. https://www.cbo.gov/publication/49891

———. 2015c. "The Effects of Potential Cuts in SNAP Spending on Households with Different Amounts of Income." Washington. https://www.cbo.gov/publication/49978

———. 2015d. "Answers to Questions for the Record Following a Hearing on the Budget and Economic Outlook for 2015 to 2025 Conducted by the Senate Committee on the Budget." Washington. https://www.cbo.gov/publication/49974

———. 2015e. "Budgetary and Economic Outcomes under Paths for Federal Revenues and Noninterest Spending Specified in the Conference Report on the 2016 Budget Resolution." Washington. https://www.cbo.gov/publication/50115

———. 2015f. "Answers to Questions for the Record Following a Hearing by the Senate Committee on the Budget on the Work of the Congressional Budget Office." Washington. https://www.cbo.gov/publication/50279

———. 2015g. "The 2015 Long-Term Budget Outlook." Washington. https://www.cbo.gov/publication/50250

———. 2015h. "Budgetary and Economic Effects of Repealing the Affordable Care Act." Washington. https://www.cbo.gov/publication/50252

———. 2015i. "Dynamic Scoring at CBO." Presentation at the NABE Foundation's 12th Annual Economic Measurement Seminar, July 21. Washington. https://www.cbo.gov/publication/50409

———. 2015j. "About CBO: Processes." Web page. https://www.cbo.gov/about/processes

———. 2015k. "Tax Relief Extension Act of 2015." Cost Estimate. Washington. https://www.cbo.gov/publication/50704

———. 2015l. "The Macroeconomic Effects of Eliminating Automatic Reductions to Discretionary Spending Caps." Letter to the Honorable Bernard Sanders, Ranking Member, Committee on the Budget, United States Senate, August 11. Washington. https://www.cbo.gov/publication/50725

Committee for a Responsible Federal Budget. 2012. "Understanding Dynamic Scoring." Washington.

Furman, Jason. 2006. "A Short Guide to Dynamic Scoring." Washington: Center on Budget and Policy Priorities.

Gale, William G. 2002. "Notes on Taxes, Growth, and Dynamic Analysis of New Legislation." *Tax Notes, 30th Anniversary Edition*: 30–43.

Hassett, Kevin A. 2002. "Dynamic Scoring." Testimony before the House Committee on the Budget, May 2. Washington. https://www.aei.org/publication/dynamic-scoring/

Holtz-Eakin, Douglas, and Michael Mandel. 2015. "Dynamic Scoring and Infrastructure Spending." New York: McGraw Hill Financial Global Institute.

Ip, Greg. 2015. "Q&A: Dynamic Scoring Is Neither Salvation Nor Scourge." Real Time Economics blog, *Wall Street Journal,* March 30.

JCT (Joint Committee on Taxation). 2003a. "Macroeconomic Analysis of H.R. 2, the 'Jobs and Growth Reconciliation Tax Act of 2003.'" *Congressional Record* 149, no. 68: H3829–H3832.

———. 2003b. "Overview of Work of the Staff of the Joint Committee on Taxation to Model the Macroeconomic Effects of Proposed Tax Legislation to Comply with House Rule XIII.3.(h)(2)." Report no. JCX-105-03. Washington.

———. 2005. "Macroeconomic Analysis of Various Proposals to Provide $500 Billion in Tax Relief." Report no. JCX-4-05. Washington.

———. 2006. "Exploring Issues in the Development of Macroeconomic Models for Use in Tax Policy Analysis." Report no. JCX-19-06. Washington.

———. 2009. "Providing for a Portion of the Economic Recovery Package Relating to Revenue Measures, Unemployment, and Health." House of Representatives Report no. 111-8, Part 1. Washington.

———. 2011a. "Summary of Economic Models and Estimating Practices of the Staff of the Joint Committee on Taxation." Report no. JCX-46-11. Washington.

———. 2011b. "Testimony of the Staff of the Joint Committee on Taxation before the House Committee on Ways and Means regarding Economic Modeling." Report no. JCX-48-11. Washington.

———. 2014a. "Distributional Effects of the 'Tax Reform Act of 2014.'" Report no. JCX-21-14. Washington.

———. 2014b. "Estimated Revenue Effects of the 'Tax Reform Act of 2014.'" Report no. JCX-20-14. Washington.

———. 2014c. "Macroeconomic Analysis of the 'Tax Reform Act of 2014.'" Report no. JCX-22-14. Washington.

———. 2015a. "Macroeconomic Analysis at the Joint Committee on Taxation and the Mechanics of Its Implementation." Presentation at the Brookings Institution, January 26. Report no. JCX-3-15. Washington.

———. 2015b. "A Report to the Congressional Budget Office of the Macroeconomic Effects of the 'Tax Relief Extension Act of 2015,' as Ordered to Be Reported by the Senate Committee on Finance." Report no. JCX-107-15. Washington.

Manski, Charles F. 2011. "Policy Analysis with Incredible Certitude." *Economic Journal* 121, no. 554: F261–F289.

Orszag, Peter R. 2002. "Macroeconomic Implications of Federal Budget Proposals and the Scoring Process." Testimony before the Subcommittee on Legislative and Budget Process, House Rules Committee, May 2. Washington. http://www.brookings.edu/~/media/Research/Files/Testimony/2002/5/02federalbudget%20orszag/20020502.pdf

Van de Water, Paul N., and Chye-Ching Huang. 2014. "Budget and Tax Plans Should Not Rely on 'Dynamic Scoring.'" Washington: Center on Budget and Policy Priorities.

Comments and Discussion

COMMENT BY

GLENN HUBBARD "Dynamic scoring" is surely an in-the-budget-weeds topic for most people, even most economists. Nevertheless, it is an important topic for study and reflection, particularly given the major tax and expenditure proposals likely to surface in the context of the 2016 presidential campaign. Answering the question of whether dynamic scoring can and should be done is straightforward, but I can attest from both government and academic experience that the ratio of heat to light in such a discussion is often high. Enter Douglas Elmendorf's paper. Elmendorf is the right author, with background both as a scholar in related research and as a distinguished former director of the Congressional Budget Office (CBO). Wisely, his paper quickly shifts the analytical discussion from "yes or no" to "how and when."

Responsible budget analysis largely incorporates estimates of the revenue or expenditure consequences of budget proposals—"scoring" them. Consider, for example, a proposal to reduce marginal tax rates across the board on individual incomes. A conventional revenue estimate would calculate lost revenue from the existing tax base and analyze the behavioral effects of the policy (for example, working harder as a result of the lower marginal tax rate on work). Such static scoring is not as naïve as the term *static* suggests; analysis of behavioral effects can be complex and rich, using a variety of elasticity estimates and microsimulation models.

While potentially rich, such work is necessarily incomplete for major proposals, in that a static analysis holds GDP constant. That is, while estimates consider compositional effects, they abstract from macroeconomic impacts. For example, suppose the Ways and Means Committee were considering a major tax reform proposal to scrap the present federal income tax on corporations and individuals and replace it with a broad-based con-

sumption tax. Almost surely, much of the motivation for such a proposal would be the goal of raising GDP and household incomes, and higher output and incomes, all else equal, will raise revenue. So conventional static scoring cannot provide the answer on *Jeopardy!* to the question, "What is the revenue impact of the tax reform proposal?"

PROS AND CONS OF DYNAMIC SCORING A consideration of the macroeconomic impacts of certain proposals is obviously the right answer to the last question. By "certain proposals" I mean to include not only tax reform but also immigration reform (with its effects on labor supply), health care reform (with effects on labor supply), and large expenditure programs (such as major infrastructure expansion initiatives). Importantly, dynamic scoring sheds light on aggregate effects that motivate congressional and White House interests in the first place. And dynamic scoring leans against the concern that current estimation procedures are stacked against policies that would advance economic growth or higher overall incomes. For example, the revenue consequences of a major tax reform would be affected according to whether the reform raised aggregate economic activity or incomes and, hence, expected revenue. In that sense, ignoring aggregate feedback effects, which are the basic element of dynamic scoring, makes the budget cost of expansionary tax and spending policies appear too expensive.

The foregoing comments notwithstanding, dynamic scoring has remained controversial in some policy circles. One objection is that formal incorporation of dynamic scoring within budget estimates introduces a bias for tax cuts—or spending increases—eroding budget disciplines. A second fear is that uncertainty in forecasts makes dynamic scoring unreliable. Finally, a technical concern is often expressed that the inclusion of macroeconomic feedback effects complicates the estimation process so much that dynamic scoring, however meritorious in theory, is simply too difficult to implement in practice. As Elmendorf observes in the paper, such concerns are off the mark.

A CONSUMPTION TAX EXAMPLE Many economists have estimated large gains on output and incomes from a shift to a broad-based consumption tax (because of reduced capital taxation and inter-asset tax distribution), including studies by academic researchers, the Treasury Department's Office of Tax Analysis, and the Joint Committee on Taxation (JCT). A study by staff economists of the JCT (1997) drew on a range of models to estimate that a shift to a consumption tax would raise GDP in steady state by 5 percent. Even the more modest reform plan in 2014 of then-House Ways and Means Committee chairman Dave Camp was estimated

by John Diamond and George Zodrow (2014) to raise GDP by as much as 3.1 percent in the long run.

With reasonable estimates of the marginal revenue effect of the change, the revenue impact would be about 1 percent of GDP in steady state, a very large adjustment. As work by Greg Mankiw and Matthew Weinzierl (2006) shows, static revenue estimates are considerably off the mark. Using a Ramsey growth model, Mankiw and Weinzierl (2006) estimate that the dynamic revenue effect of a cut in the capital income tax rate is about half the static estimate. (The difference would be modestly attenuated with a finite-horizon case and accentuated in a model with imperfect competition and markups.)

Again, as Elmendorf notes, there are legitimate concerns that macroeconomic modeling is not an exact science—but the same can be said for microsimulation models underlying conventional analysis. In addition, some fears about dynamic scoring are really concerns about the distributional consequences of proposals such as tax reform—of course, information on both budget and distributional effects should be presented to policymakers. Finally, the question arises as to whether dynamic scoring should be incorporated in budget rules—but surely such information is the best answer for major proposals.

Returning to the consumption tax example: failure to provide dynamic scoring denies policymakers information on the economic gains from tax reform. That lack of information effectively denies policymakers the ability to understand and manage trade-offs between the distributional and economic efficiency consequences of tax reform.

Finally, the uncertainty argument against dynamic scoring of tax reform proposals is not compelling. The idea that the method's uncertainty suggests the desirability of adopting a static-scoring answer that is *known with certainty to be incorrect* is not logical. And while the Federal Reserve must formulate monetary policy in the face of macro uncertainty, its own reliance on economic models is uncontroversial.

NEXT STEPS I agree with Elmendorf's basic point that the task ahead is to figure out how and when to do dynamic scoring, ending the existential debate. I think of this next step as incorporating analysis, process, and politics.

Analysis A key first step is to define candidate policies for dynamic scoring. Such candidates include policies with a material impact on aggregate demand, productivity, and/or hours worked. The second step is to encourage the staff economists of the JCT, the CBO, and the Office of Tax Analysis to refine models, including open-economy features and realistic

heterogeneity on the household side, with exercises incorporating the contribution and judgment of outside experts. Third, toward this end, organizations like the National Bureau of Economic Research and think tanks can develop forums to investigate the professional consensus regarding aggregate economic effects of major tax and spending proposals.

Process Moving to dynamic scoring requires a shift in both resources and best practices. That is, integrating microeconomic and macroeconomic models would require substantial incremental financial resources for the staffing of official scorekeepers. And a dynamic score should provide information as to why an estimated macroeconomic effect differs from a consensus estimate by economists. Assessing research on the best process would be enhanced by the use of a panel of outside experts, by official scorekeepers, and by the existence of a nonofficial, "open source" alternative that could be employed by outside researchers or policymakers seeking advice.[1]

Politics The political concern that dynamic scoring will inappropriately soften attention to the budget deficit must be addressed head on. The key, as Elmendorf notes, is to focus dynamic scoring on major proposals and to use consensus estimates of the macroeconomic effects of policy changes. In this regard, analysis of tax policy changes is more straightforward than analysis of spending changes, since it is easier to estimate the aggregate effects of tax policy over short- and medium-term horizons. More research on the spending side will help frame potential dynamic scoring of spending programs for infrastructure, education, and training support. Finally, political questions about the applicability of dynamic scoring to budget rules must be addressed. The answer here should be simple: To the extent that dynamic scoring is conceptually correct (it is), and implemented rigorously (it can be), the dynamic score should be the official score under budget rules. Including it only in an impact statement presented to decisionmakers is too limiting.

Elmendorf's thoughtful and careful paper makes a strong case for dynamic scoring and identifies next steps for implementation. Read one

1. The Open Source Policy Center, recently inaugurated by the American Enterprise Institute, is a welcome development in this regard. Incorporating dynamic scoring involves exporting output from microsimulation models of policy changes to a dynamic macroeconomic model that models a substantial portion of aggregate economic activity. In this model-bridging exercise, the macroeconomic model output from one time period can be fed into the microsimulation model for the next period. Openness of both models exposes assumptions and can allow policy analysts to pinpoint sources of disagreement.

way, the paper is so sensible and straightforward it seems "apple pie." But I think it reflects how far this policy debate has come. A generation ago, while working as a tax official at the Treasury Department, I occasionally felt like a referee on dynamic scoring between proponents who thought it was the savior of economic policy and opponents who thought it represented the end of western civilization. That we are now having a reasonable discussion about *how* to implement dynamic scoring is surely a good sign.

REFERENCES FOR THE HUBBARD COMMENT

Diamond, John W., and George R. Zodrow. 2014. "Dynamic Macroeconomic Estimates of the Effects of Chairman Camp's 2014 Tax Reform Discussion Draft." Houston: Tax Advisers LLC. http://businessroundtable.org/sites/default/files/reports/Diamond-Zodrow%20Analysis%20for%20Business%20Roundtable_Final%20for%20Release.pdf

JCT (Joint Committee on Taxation). 1997. "Joint Committee on Taxation Tax Modeling Project and 1997 Tax Symposium Papers." Report no. JCS-21-97. Washington.

Mankiw, N. Gregory, and Matthew Weinzierl. 2006. "Dynamic Scoring: A Back-of-the-Envelope Guide." *Journal of Public Economics* 90, nos. 8–9: 1415–33.

COMMENT BY

DONALD B. MARRON Douglas Elmendorf lays out a compelling case that the Congressional Budget Office (CBO) and the staff of the Joint Committee on Taxation (JCT) should account for macroeconomic feedback when scoring proposed legislation. I agree. The two agencies have been developing and refining their macroeconomic modeling techniques for more than a decade. They have successfully applied those techniques in a host of analyses. The next step in this gradual evolution is to incorporate them in official scoring.

Concerns about dynamic scoring are understandable in light of highly politicized fiscal policy debates and the way claims about dynamic effects often align with ideological views. But CBO and JCT have a strong track record of navigating such shoals and delivering nonpartisan analysis to Congress. I expect the same will be true with dynamic scoring, which will not fully live up to the hopes of its proponents nor "live down" to the fears of its detractors. Instead, it will modestly improve the budget estimates that inform policymakers and the public.

Seven points in Elmendorf's paper deserve particular emphasis:

First, including macroeconomic effects *can improve budget estimates.* The budget process requires year-by-year estimates of the revenue and spending implications of proposed legislation. CBO's and JCT's missions are to make those estimates as accurate as possible, given the constraints under which they operate. Fiscal policies can change how people work, save, invest, and spend and thus can raise or lower macroeconomic activity. Including such effects in official scores will improve budget deliberations as long as there is a sufficient evidentiary base for estimating them and as long as doing so is consonant with other constraints the agencies face (such as timeliness, resource limits, and transparency requirements).

Second, including macroeconomic effects *can improve policy comparisons.* Policy debates often distill macroeconomic effects to the level of tweets and bumper stickers: "Tax cuts boost growth"; "Spending stimulates the economy." The extent to which such claims are true, however, depends on policy specifics. Marginal tax cuts likely do more for long-run economic growth than do inframarginal ones, for example, and spending in recessions likely boosts the economy more than at times of full employment (CBO 2014).

CBO and JCT have traditionally quantified those differences in supplementary analyses undertaken outside of official budget scoring. Including them in official scores would make them more politically salient. Tax cuts that weaken long-run growth, for example, will get worse budget scores than under conventional estimating, while cuts that encourage growth will get better scores. Differing scores will give lawmakers more reason to consider seriously the macroeconomic effects of competing policy proposals.

Third, dynamic scoring *should apply to both taxes and spending.* Most of the public debate about dynamic scoring has focused on tax policies, but spending programs have the same potential for macroeconomic effects. Investments in infrastructure and education can boost long-run economic potential; so can private investment induced by tax policy changes. The phase-out of benefits in social insurance programs can discourage labor supply; so can taxes on wages. Spending can soften recessions; so can tax cuts. Spending financed by deficits can crowd out private investment; so can tax cuts financed by deficits. Treating spending and taxes equally in dynamic scoring thus makes perfect sense. This is straightforward for taxes and mandatory spending, but it is more challenging for the discretionary spending that Congress handles through its annual appropriation process. Whether and how to apply dynamic scoring to discretionary spending therefore deserves further attention in budget process discussions.

Fourth, dynamic scoring *should be applied only to major pieces of legislation.* Dynamic scoring is logistically challenging. It takes time and talented staff. As a result, there are practical limits to how many dynamic scores CBO and JCT can produce. It makes sense to focus solely on the largest bills, while allowing legislators to require dynamic scoring in other cases where macroeconomic effects may be important. If CBO and JCT develop ways to make macroeconomic analysis easier in the future, lawmakers can expand the scope of required dynamic scoring.

Fifth, *the authority to require dynamic scoring should be shared by the two parties.* Under current congressional procedures, only the chairs of the House and Senate budget committees can request dynamic scoring of particular bills. As Elmendorf recommends, however, that power should also be given to the ranking members of those committees. Sharing this power would reduce the risk of the majority's using dynamic scoring strategically and would maintain more consistency in the scoring's application.

Sixth, the adoption of dynamic scoring at the start of 2015, as important as it is, *is not as big a break with past practice as it first appears.* CBO and JCT have been publishing dynamic analyses for more than a decade, including analyses of major tax reforms (JCT 2014a, 2014b), stimulus proposals (CBO 2009), and presidential budgets (CBO 2015). The techniques used in those analyses have been refined through experience and external review.

Immigration reform proposals in 2006, 2007, and 2013 provided special opportunity for the agencies and Congress to prepare and consider scores that included some macroeconomic effects. When CBO and JCT did this in 2006—at a time when I served as CBO's acting director—we expected controversy, but there was none; lawmakers and outside analysts understood that it made sense for CBO and JCT to consider the effects of an increased labor force when evaluating immigration reform despite the convention of not including any macroeconomic effects for other bills (Marron 2013).

In addition, the agencies incorporate the macroeconomic effects of fiscal policy when constructing their twice-yearly budget baselines. In August 2012, for example, CBO had to project the budget outlook in the face of the then-looming "fiscal cliff," a panoply of scheduled tax increases and spending cuts. Those provisions would have amounted to $500 billion in fiscal tightening in 2013. As a result, CBO (2012) projected that the economy would be pushed into recession. That projection provided important context for fiscal cliff deliberations.

Dynamic analysis, immigration reform, and baseline projections have thus paved the way for the agencies to include macroeconomic feedbacks in official scores.

Seventh, *the strongest concerns about dynamic scoring—uncertainty and the risk of bias—are understandable but manageable.* Macroeconomic impacts are uncertain, and experts disagree on how best to model them. How myopic or forward-looking are individuals when making work, saving, and consumption decisions? How will the Federal Reserve respond to changes in fiscal policy? What are the feedbacks between the U.S. economy and the rest of the world? How much do deficits crowd out private investment? Can fiscal policy reduce hysteresis effects in the aftermath of a deep recession?

In principle, such uncertainties could create opportunities for the agencies to put a thumb on the scale to favor results preferred by their political masters. In reality, the cultures and staffing of CBO and JCT are fundamentally nonpartisan. The two agencies' directors are chosen by the congressional majority, and they certainly hear from congressional leaders about important bills, but they have done an admirable job maintaining their nonpartisan credibility.

Moreover, concerns about uncertainty and potential bias apply equally to many estimates the agencies have traditionally produced. Future economic conditions and behavioral responses are highly uncertain. What will be the take-up rate of a never-before-seen benefit program? How will state governors and legislatures respond to new flexibility in a federal program? What are the probability and magnitudes of potential terrorist attacks in the United States? What will oil prices be in 2025? What will electromagnetic spectrum sell for in 2022? What new medicines will come to market in the next decade? At what price and usage? How will the Supreme Court respond to a possibly unconstitutional piece of legislation?

The strong nonpartisan culture of the agencies has allowed them to make objective calls on such questions in the past and will allow them to do so with dynamic scoring in the future. That culture also implies that the effects of dynamic scoring will be less rosy than some proponents claim and less dire than some opponents fear. President Harry S. Truman once asked for a one-armed economist. He would not find any at CBO and JCT. Instead, the agencies will provide classic many-handed analyses that include offsetting effects.

In analyzing macroeconomic feedback from tax cuts, for example, the agencies will consider how they might encourage working, saving, and

investing. But they will also consider how tax cuts may increase after-tax income and thus reduce work and how any resulting deficits may eventually reduce private investment. To use the jargon, CBO and JCT will consider the income effects and the crowding-out effects of tax cuts, not just the substitution effects that proponents emphasize. That three-handed approach tempers the potential macroeconomic effects of tax cuts. Indeed, it reveals that some tax cuts reduce economic growth and thus have a larger budget cost than conventionally estimated (CBO 2010).

The same is true of the dynamic scoring of spending provisions, stimulus efforts, and other policies that often have the opposite political valence. Proponents of stimulus often emphasize the potential boost from putting money in peoples' pockets and the multiplier effects that this may set in motion. But CBO and JCT also consider whether and how much Federal Reserve policy may offset such effects and what long-term drag will result from accompanying deficits and accumulated debt (CBO 2014).

The reality of dynamic scoring is thus unlikely to live up to the hype. Instead, dynamic scoring will modestly improve the budget projections that inform fiscal policy deliberations.

REFERENCES FOR THE MARRON COMMENT

CBO (Congressional Budget Office). 2009. "The State of the Economy and Issues in Developing an Effective Policy Response." Statement of Douglas W. Elmendorf before the Committee on the Budget, U.S. House of Representatives, January 27. Washington. https://www.cbo.gov/publication/41755

———. 2010. "The Economic Outlook and Fiscal Policy Choices." Statement of Douglas W. Elmendorf before the Senate Committee on the Budget, September 28. Washington. https://www.cbo.gov/publication/21836

———. 2012. "An Update to the Budget and Economic Outlook: Fiscal Years 2012 to 2022." Washington. https://www.cbo.gov/publication/43539

———. 2014. "How CBO Analyzes the Effects of Changes in Federal Fiscal Policies on the Economy." Washington. https://www.cbo.gov/publication/49494

———. 2015. "A Macroeconomic Analysis of the President's 2016 Budget." Washington. https://www.cbo.gov/publication/50734

JCT (Joint Committee on Taxation). 2014a. "Estimated Revenue Effects of the 'Tax Reform Act of 2014.'" Report no. JCX-20-14. Washington.

———. 2014b. "Macroeconomic Analysis of the 'Tax Reform Act of 2014.'" Report no. JCX-22-14. Washington.

Marron, Donald. 2013. "Immigration, Dynamic Scoring, and CBO." Blog post, May 3. Washington: Urban-Brookings Tax Policy Center, TaxVox. http://taxvox.taxpolicycenter.org/2013/05/03/immigration-dynamic-scoring-and-cbo/

GENERAL DISCUSSION Peter Orszag spoke first to say that the arguments against dynamic scoring have never been theoretical but, instead, have been pragmatic. A political economy argument can be made on both sides. Moreover, the Congressional Budget Office (CBO) has been providing estimates of macroeconomic feedback effects for some time in the form of dynamic budget analysis. The question is whether they should be directly incorporated into the budget score. Echoing discussant Donald Marron—who worked as acting director of the Congressional Budget Office (CBO) in 2006—Orszag admitted that a decade ago he would have been more concerned about dynamic scoring. Nowadays he is less concerned, and he agreed with Douglas Elmendorf that it is marginally better to incorporate dynamic scoring directly into the budget score than not to. In his view, the debate over the 2009 stimulus package would have been much better informed if this had been done at the time.

Orszag offered two notes of caution. First, he stressed the importance of applying dynamic scoring to both spending *and* tax proposals, noting that this may not necessarily be consistent with the current political environment. Second, he warned about potential abuses of the process, and offered two suggestions to protect against it. The appointment of the CBO director has traditionally been somewhat bipartisan, with informal cooperation between the chairs and ranking members, but that arrangement can easily fall apart, especially in times like the present where polarization reigns. The answer is to make the appointment formally bipartisan. Orszag also would encourage outside entities—such as the Tax Policy Center—to play a greater role than they do now in providing a check on the reality of what the CBO is doing in the dynamic scoring.

Alan Blinder agreed with Orszag that it is important to think about how to increase the independence of the CBO from political meddling given the new scope for difficult judgments that dynamic scoring would open. He pointed to the well-known fact that long-term projections, which often go "out of sample," can magnify standard errors, which makes it all the more important to protect the neutrality of the estimating process. Blinder suggested that the greater independence from political meddling enjoyed by the Federal Reserve Board is a standard that should be aimed for, hard as that may be.

William Gale voiced similar concern over the political economy implications. He argued that one could conclude from Elmendorf's paper that while dynamic scoring should be used, the way the House currently narrows its application to tax cuts and disallows the minority from requesting it renders it unbalanced. Gale felt that as a budget rule, dynamic scoring

could be used to prop up political agendas as much as it could be used for illumination.

Alan Viard also acknowledged that the risks of political bias could be heightened by requiring dynamic scoring. However, he added, estimates of macroeconomic effects by a respected agency whose directors have come from both political parties could help combat exaggerated claims presented by outside advocates, such as claims that tax cuts are likely to fully pay for themselves.

Jason Furman pointed to the value of the second half of the paper, which discussed the CBO's role in providing other information to policymakers that may be relevant to them, in addition to estimating budget impacts. This role deserves more careful thought, since some policymakers can have limited attention spans and often ignore important details of interpretation. For example, fiscal policy analysis that just shows growth or jobs effect but gives short shrift to welfare effects reflects an elementary cost-benefit mistake we would not make in the regulatory arena. Furman stressed that growth effects should be reported in tandem with welfare effects, since policymakers will not make the right choices if CBO only reports growth or job metrics without embedding them into a broader context of distribution and welfare. He argued for the creation of some type of simple summary statistic—not just in footnotes or caveats—to capture such total effects.

Gregory Mankiw characterized Elmendorf's paper as eminently sensible and then also raised two points of concern. He noted that in the federal government, a lot of attention is paid to the 10-year budget window, and the CBO projections are no different. From a president's standpoint, 10 years may seem like forever, but for the economy it is a relatively short time frame. Mankiw's worry was that the focus on a 10-year frame might overemphasize short-run Keynesian demand effects and underemphasize longer-run classical growth effects. Dynamic scoring—which currently projects only 10 years into the future—could also lead to shortsightedness. He recommended that a balance be struck in the estimating between short-run and long-run effects.

Along similar lines, Mankiw also stressed that any well-specified general equilibrium model must close budget gaps in the long run, otherwise the proposal will be incomplete. Proposals that cut taxes without stating how the budget gap will be closed in the future are incomplete, and so are those that propose to increase spending without providing a way to pay for it. Outcomes will depend crucially on how the policies are closed. To take two current examples, one might propose to pay for a stimulus package by cutting Social Security in the long run; or by increasing capital taxes

in the long run. In fact, those two approaches would have very different steady-state effects. On this point, he strongly disagreed with Elmendorf, who seemed to say that if one had a well specified model, how this loop was closed would not matter for the long-run effects. In short, Mankiw felt, dynamic scoring needs to take into account the long-run effects of closing policy proposals.

Gale echoed the importance of fiscal closure rules. He agreed with Mankiw on the principle that tax cuts financed by future tax increases have different long-term consequences than future spending cuts. Budget constraints are real, and policymakers must specify how a policy is going to be paid for.

Martin Feldstein praised Elmendorf's paper for its advocacy of dynamic scoring as well as its discussion of the technical problems associated with it. However, he expressed worry in how dynamic scoring is actually implemented. He reminisced about the "bad old days" of budget scoring, before the mid-1970s, when revenue estimation was done with the assumption that changes in tax rates had no effect at all on taxpayer behavior. This assumption was even applied to taxation of capital gains, which Feldstein noted can actually have a very large effect. He recalled how when he explained this assumption in testimony before the Senate Finance Committee in the late 1970s, Senator Russell Long's shocked reaction shamed the estimators into changing their approach. Since then the joint tax committee and the Treasury staff have applied the convention that changes in tax rates do change behavior, though not in the way Elmendorf has emphasized, with an impact on GDP.

In thinking about the response to changes in the personal income tax, Feldstein distinguished between three kinds of effects, which he believed were not adequately captured by the current dynamic scoring methods. The first kind are the short-run aggregate demand effects, including how the Federal Reserve responds to offset the fiscal impact. The second are the longer-run growth effects with their impacts on savings, investment, and human capital. The third kind—which in Feldstein's view are the most overlooked—are the permanent revenue effects of behavioral changes in response to changes in marginal tax rates. To illustrate, he named three ways behavior changes in response to a reduction in the marginal tax rate: by increasing the labor supply, broadly defined, including occupational choice; by influencing the form of compensation, since workers will prefer taxable cash over benefits when rates are lowered; and through changes in spending on tax-favored consumption, such as mortgage interest and charitable contributions. He said the response of labor supply elasticities,

hours, and participation to tax rate changes is small in comparison with the response of taxable income, according to available microdata. Any dynamic model needs to look much more closely at the latter.

Jeffrey Kling, representing the CBO, assured Feldstein that the CBO has long attempted to incorporate some of the issues he described regarding compensation and the labor supply. Its staff continues to study, including impacts on taxable income, to try to synthesize the findings and apply them to proposals that would affect forms of compensation. At the same time, he noted that while the CBO could probably do more in this area, it actually falls under the jurisdiction of the Joint Committee on Taxation.

Alice Rivlin noted that most of the discussion thus far had been about tax changes, with little attention given to government spending and its macroeconomic effects. She acknowledged that spending is much harder to model, both because the evidence base is not nearly as well developed as for taxation and because many public investments, such as those in early childhood education, can only improve productivity over a very long term and in the aggregate. However, Rivlin stressed, it is as important to consider the very long-range effects of spending as it is to examine the very long-run effects of taxation.

Caroline Hoxby noted that to a microeconomist, taking the central tendency of estimates seems quite unnatural. Microeconomists would prefer to take the best econometrically identified estimate, the one that is closest to what one believes would occur were a policy to go into effect, ideally based on using a randomized trial. Applying a midpoint or consensus estimate essentially assumes that mistakes made on one side are offset by those made on the other or that modeling decisions somehow are distributed in some normal way.

Kling—again speaking for the CBO—clarified that a "central estimate" is informed by a judgment of the reading of the literature, which is not necessarily arraying all of the point estimates and taking the midpoint. As an example, if analysis showed that a policy had a 60 percent chance of having no effect and a 40 percent chance of having a positive effect, the CBO would synthesize that as some positive number.

Justin Wolfers remarked that there seemed to be an emerging consensus in the room that there need be no more debate about the principle of dynamic scoring. If so, he disagreed. One reason to continue the argument over its merits stems, he said, from a simple statistical principle of shrinkage estimators. When one has a raw, unbiased but noisy statistical estimate, a forecast is improved in a mean-squared-error sense by shrinking it back to some prior, yet for many of the policies under discussion the prior

was simply that "anything could happen." If that is the case, according to Wolfers, the right thing to do would be to shrink toward zero. This led him to agree with Marron, who proposed that the midpoint of macro thinking indeed adds value. He concluded that fighting about dynamic scoring is sort of a second-best way of actually getting shrinkage estimators and more efficient estimates.

Wolfers also returned to the earlier points made about political bias. Will dynamic scoring yield a pro–tax cut bias? He proposed thinking about the social and political leanings of people within the economics profession itself. One could caricature (in good humor) the profession as consisting of people who hate inflation becoming monetary economists, those who love global trade becoming international economists, those who like workers becoming labor economists, and those who like capital and rich people becoming financial economists. Wolfers conjectured that those in the field of public finance—and particularly those who analyze tax cuts—contain a higher share of Republicans than any other fields in the profession, and if that really is so it could add a pro–tax cut bias to their dynamic scoring.

Brad DeLong noted that when he was a Treasury political appointee, one of the Treasury career staff economists lectured him about dynamic scoring thus: "Brad, you people come in with your exaggerated belief in the productivity benefits of public investment. And so you command us to score your policies as having a very favorable impact on the deficit. They come in with their exaggerated belief in the benefits of tax cuts. They command us to score their policies as having a very favorable impact. We cannot say we disagree with our bosses' analytic judgments. But by holding the line and stating that we do not consider any macroeconomic effects of policies, we can at least prevent being whipsawed by this partisan rosy-scenario ratchet."

He noted that being whipsawed by the partisan rosy-scenario ratchet is a serious danger, as evidenced most recently by the recent semi-score of the Jeb Bush tax plan by Feldstein and others.[1] There would be an upside if appropriate real technocratic dynamic-scoring corrections were significant. But, he concluded, they mostly likely are not.

Phillip Swagel raised the issue of Congress's budget process itself and its influence on policies and the economy. Economists too often overlook the significance of budget rules, he said, including Senate procedures, which

1. John Cogan, Martin Feldstein, Glenn Hubbard, and Kevin Warsh, "Fundamental Tax Reform: An Essential Pillar of Economic Growth; An Assessment of Governor Jeb Bush's 'Reform and Growth Act of 2017'" (New York: Center for Global Enterprise, 2015).

have a real effect on economic outcomes. Swagel has been impressed by the CBO's and the Joint Committee on Taxation's willingness to push back against budget gimmicks, citing as an example their analysis of proposals for repatriation of foreign earnings.

Viard voiced support for dynamic scoring, noting that its most important role was not to favor or disfavor tax cuts, as Wolfers had suggested. Rather, he suggested that a more important purpose was to help illuminate which kinds of tax cuts are better for economic growth than others. Similarly, the extension of dynamic scoring to spending proposals could help illuminate which kinds of spending are better for growth.

George Perry voiced a concern that no one else had raised, namely that while it had been discovered fifty years earlier that fiscal policy has a special role to play in a depressed economy as a stabilization tool, today this gets overlooked in a scramble to tease out what he considers third-order effects. Supply-side effects, like hysteresis and the use of unemployment insurance are, in the bigger picture, not nearly as important in an underemployed economy as the first-order effects of fiscal policy.

Elmendorf spoke in response to all the comments, first by agreeing with Furman about the importance of distinguishing between the effects on GDP and the effects on people's welfare. He noted that in its analysis of immigration reform, the CBO was careful to distinguish between total GDP and GDP per capita, which is a better measure of welfare. He noted that it is always a challenge when publishing analyses to explain what the numbers mean, since so many policymakers will go just to the numbers without reading the words around them.

Elmendorf disagreed with Mankiw's view that too much attention to the 10-year window made it harder to properly understand the steady state. He believes the CBO's analyses of short-run effects of policy changes remain vital, as they can be very large effects. And because policies are rarely permanent, whether the long-run steady-state effects of a policy can ever be realized is unclear. Moreover, when the CBO does think the longer-term effects of a policy will differ from the effects in the first decade, it generally explains those longer-term effects as well, its work on the Affordable Care Act and immigration reform being two salient examples.

To the question about the value of picking a midpoint estimate in dynamic modeling, he answered that in his first few years at the CBO, the practice was to publish only the low end and the high end of the range. That left policymakers free to pick their favorite number at the extreme, however, and offered them no other simple choice. The CBO

improved on this a few years ago by adding in the central number and releasing that along with the full range.

Addressing the political economy aspect of dynamic scoring and the independence of the CBO and the staff of the Joint Committee on Taxation, Elmendorf assured everyone that there is an ongoing understanding and respect among congressional leaders for that independence. He pointed out that he had originally been appointed director of the CBO by two Democrats and was reappointed by a Democrat and a Republican. Likewise, Thomas Barthold was originally appointed as the staff director for the Joint Committee on Taxation when Democrats controlled both the Finance and Ways and Means committees, and was subsequently reappointed when Republicans controlled both. In response to Blinder's comparison with the greater independence of the Federal Reserve Board, he noted that many Members of Congress are in fact not very happy with having given so much latitude to the Federal Reserve, and surely they would be much less happy doing the same for agencies working on fiscal policy.

SATEN KUMAR
Auckland University of Technology

OLIVIER COIBION
University of Texas, Austin

HASSAN AFROUZI
University of Texas, Austin

YURIY GORODNICHENKO
University of California, Berkeley

Inflation Targeting Does Not Anchor Inflation Expectations: Evidence from Firms in New Zealand

ABSTRACT Using a new survey of firm managers, we investigate whether inflation expectations in New Zealand are anchored or not. In spite of 25 years of inflation targeting by the Reserve Bank of New Zealand, firm managers display little anchoring of such expectations. We document this finding along a number of dimensions. Managers are unaware of the identities of central bankers or of central banks' objectives, and they are generally poorly informed about recent inflation dynamics. Their forecasts of future inflation reflect high levels of uncertainty and are extremely dispersed, and they are volatile along both short-run and long-run horizons. Similar results can be found for the United States using currently available surveys.

> *The effects of monetary policy on the economy today depend importantly not only on current policy actions, but also on the public's expectations of how policy will evolve. . . . Indeed, expectations matter so much that a central bank may be able to help make policy more effective by working to shape those expectations.*
>
> —Ben Bernanke (2013)

Central bankers increasingly emphasize the importance of the public's expectations. One reason is that unanchored inflation expectations are commonly viewed (Solow 1979) as having played an important role in the Great Inflation of the 1970s and the subsequent large costs of bringing down inflation over the course of the 1980s. Maintaining low and stable "well-anchored" inflation expectations has become a mantra of modern central banking. But with the onset of the zero bound on interest rates, expectations have also taken a new role as a potential instrument of

monetary policy. By trying to raise inflation expectations when they are very low, central bankers can immediately lower real interest rates and thereby stimulate economic activity even when nominal rates are constrained, a strategy actively pursued by the Bank of Japan, for example.

In this paper, we investigate both the question of whether inflation expectations are indeed well "anchored" and whether monetary policies designed to influence inflation expectations are likely to be successful. To do so, we rely primarily on a recent survey of managers of firms in New Zealand, the country that pioneered inflation targeting in 1989. In this survey, we asked managers a wide range of questions about their inflation expectations and their individual and firm's characteristics as well as their knowledge and understanding of monetary policy. Following Carola Binder (2015), we argue that many of our results extend to the United States using existing survey data. The New Zealand survey fills an important gap in our understanding of expectations. Previously, the only quantitative macroeconomic surveys available were those covering professional forecasters, very large firms, or households. The first two categories of respondents are, obviously, extremely well informed, but the latter tend to be poorly informed. Our survey focuses on decisionmakers within a wide range of firms and therefore provides a much-needed middle ground for assessing the economic knowledge and understanding of individuals who make pricing, hiring, and investment decisions in the economy.

Our results are not favorable to policymakers. Despite 25 years of inflation targeting in New Zealand, managers of firms there have been forecasting much higher levels of inflation than has actually occurred, at both short-run horizons and very long-run horizons. Their average perception of recent inflation is also systematically much higher than actual inflation. There is tremendous disagreement in forecasts among managers, at all horizons, as well as disagreement about recent inflation dynamics. Managers also express far more uncertainty in their inflation forecasts than do professional forecasters.

Other characteristics of firms' forecasts are also at odds with even weaker definitions of "anchored" expectations. Because managers were surveyed on multiple occasions, one can consider the revisions in their forecasts. One would expect individuals whose forecasts are anchored to display only small revisions in the forecasts they make over time. Instead, we find that managers commonly report large revisions in their forecasts. Similarly, we find that managers who expect high inflation in the short run also tend to expect higher inflation in the long run, whereas the anchoring of expectations around a well-known target should imply little (or at

least much less) co-movement between short-run inflation expectations and longer-run inflation expectations. We show formally that each of these results corresponds to different definitions of anchored expectations, but our data from managers' inflation expectations systematically reject *all* definitions of anchoring that we consider.

How could 25 years of inflation targeting have so little effect on managers' inflation expectations? We find little evidence that managers question the credibility of the Reserve Bank of New Zealand (RBNZ): the vast majority of respondents believe that the central bank can control inflation at both long-run and short-run horizons. However, many respondents display surprisingly little knowledge about the institution itself. When asked who the governor of the RBNZ is, only 30 percent chose the correct answer (out of four possible answers). Very few respondents knew that the RBNZ has inflation targeting in its mandate, and even fewer could name the specific inflation target of the RBNZ. All of this suggests that knowledge about the RBNZ and its objectives remains very limited in New Zealand, even among firm managers.

Those managers who know more about the objectives of the RBNZ have much better information about recent inflation dynamics, make forecasts that are much closer to those of professional forecasters, and have less uncertainty in their inflation forecasts than others. They follow the news more closely than do other managers and report that the news is more important to their business decisions. Significantly, these managers also report that they would be willing to pay much more for inflation forecasts (both in nominal terms and as a share of their firm's sales) than would managers who display less knowledge about monetary policy. This suggests that managers differ in their perceptions of the value of acquiring information about inflation and that these differences are reflected ex post in their knowledge of both the central bank's actions and actual inflation dynamics.

What might explain these differences? Not surprisingly, managers who are more educated have systematically better knowledge of monetary policy. More interestingly, we find that some firm characteristics help predict a manager's knowledge of monetary policy in ways that are intuitive once interpreted in a rational inattention perspective (Sims 2003). For example, when firms face more competitors, managers have more incentive to collect information about economic conditions and tend to make smaller errors about monetary policy. When firms sell a larger share of their products outside of New Zealand, managers have less incentive to track conditions in New Zealand and make larger errors about monetary policy in New Zealand. Olivier Coibion, Yuriy Gorodnichenko, and Saten

Kumar (2015) similarly find that rational inattention motives can account for a nontrivial share of the size of errors that firm managers make about recent inflation dynamics.

While these observable firm characteristics clearly play an important role in accounting for differences in managers' knowledge about monetary policy and inflation, much of the heterogeneity in beliefs remains unexplained. To get at these deeper underlying differences, we conducted a smaller wave of the survey to extract narrative responses from managers about how they form and use their inflation expectations. These responses showed that differences in managers' inflation forecasts and perceptions do not reflect confusion about what inflation means. When asked to define inflation in open-ended questions, almost 90 percent do so correctly. Nor are managers prey to conspiracy theories about statistical agencies misrepresenting inflation numbers: The vast majority report that they believe that statistical agencies correctly measure inflation. They also agree to a surprising extent about inflation rates at the level of individual categories of goods. In fact, they agree *more* with each other about recent price changes for most categories of goods than they do about recent aggregate price changes. We document that the wide disagreement about the latter is instead driven primarily by disagreement among managers about the relative importance of different categories in constructing price indexes (such as weights on house prices versus food prices versus gasoline prices).

We also asked managers about how they form their inflation expectations and what sources of information they use. In open-ended questions, most managers responded (in almost equal proportion) either that they relied on the media or that they relied on their personal shopping experience to inform them about prices. Those for whom the media was the primary source made smaller errors, on average, when asked about recent inflation dynamics. We also asked managers to quantitatively rank a wider set of information sources about inflation. This revealed that while only 20 percent of managers relied on professional forecasts, they had by far the best information about inflation.

Perhaps most strikingly, almost 90 percent of managers rated their personal shopping experience as very or extremely important to them in informing them about inflation, and 76 percent rated gasoline prices the same way. Hence, the vast majority of managers, even those who also follow newspapers and professional forecasts, report that their personal exposure to the individual prices they face (and gasoline prices in particular) plays a large role in accounting for their inflation expectations. The majority of managers also report that the primary use of their inflation expectations is for their personal shopping decisions rather than in

business decisions. The fact that most managers cite their personal shopping experience as both the primary source and the primary use of their inflation knowledge provides one rationale for why managers' expectations resemble those of households and why they disagree so much about the relevant weights to apply to different categories of goods in aggregating across them.

This is not to say that inflation expectations are irrelevant to managers' business decisions. When asked if higher inflation expectations on their part would specifically affect any of their business decisions (such as pricing and wages), most managers report that they would change some element of their business. Coibion, Gorodnichenko, and Kumar (2015) found in an experiment that when managers were presented with information about the central bank's inflation target, they tended to significantly revise their inflation forecasts toward the target, especially if they were initially very uncertain about their forecast. Jointly, these results suggest that *if* central banks could more successfully communicate their objectives to the firm managers, it would have repercussions on managers' economic decisions. The challenge for monetary policymakers, however, is that most managers currently appear to devote very little effort to tracking macroeconomic information, much less information from the central bank. Very few ever read monetary policy reports, receive Twitter feeds from the RBNZ, or receive other forms of direct communication used by the RBNZ. This absence of even basic knowledge about the central bank of New Zealand on the part of business leaders suggests that monetary policies designed to operate through changes in the public's expectations, as induced primarily through communications policies, are unlikely to be very successful under current conditions.

New Zealand, because of its exceptionally long history of inflation targeting and stable inflation, is a particularly apt place to stage a survey to gauge the extent to which anchoring of expectations can be achieved. Our finding that the inflation expectations of business leaders are no more anchored than those of households in New Zealand is therefore particularly striking, as is the finding that few business leaders are even slightly knowledgeable about monetary policy. But as documented in Binder (2015) and further extended here, many of our results carry over to the United States, at least for the broader population. Using the University of Michigan's Survey of Consumers and the Federal Reserve Bank of New York's Survey of Consumer Expectations, we find all the same patterns in inflation expectations as we previously documented for managers of firms (as well as households) in New Zealand. Expectations in the United States, therefore, appear to be just as unanchored as they appear to be in New Zealand.

In addition, polling data similarly confirm that the public in the United States shows no more knowledge of monetary policy than the public in New Zealand. Americans have great difficulty in identifying the chair of the Federal Reserve Board of Governors and are generally unable to identify recent inflation dynamics with any degree of precision. When asked about inflation over 10 years, few are willing to confidently predict low levels of inflation, a finding that speaks either to the low credibility of the Federal Reserve or, more likely, to the fact that most people do not know what reasonable ranges of inflation rates are. Nor do they seem to show much interest in learning about monetary policy. Twitter and Facebook followers of the entire Federal Reserve System are outnumbered by followers of the FBI and the CIA, and they barely outnumber the followers of former congressman Ron Paul or Senator Rand Paul. Paul Krugman singlehandedly has almost twice as many Twitter followers as the entire Federal Reserve System. Google searches confirm this paucity of interest: Online searches for macroeconomic variables like GDP, unemployment rate, and inflation are consistently topped altogether by online searches for puppies.

This paper builds on a growing literature on central bank policies and communication and their effects on expectations, yielding mixed evidence on the degree to which inflation targeting anchors expectations (for example, see the recent survey by Alan Blinder and others [2008]). However, this literature has focused almost exclusively on the expectations of financial markets or professional forecasters, primarily due to data limitations. One particularly remarkable exception is Binder (2015), who uses existing survey and polling data to assess what the U.S. public knows about monetary policy and on whose work we build explicitly. One implication of our results, along with Binder's, is that future theoretical work should focus on models in which different types of agents, such as financial market participants, consumers, and managers, form their expectations differently from one another. Only with this type of model will we be able to fully understand how large the potential benefits might be from anchoring the expectations of consumers and managers.

Also particularly relevant to our work is the seminal firm-level survey of Blinder and others (1998). We closely follow the approach they pioneered, in the sense that we aim to build a nationally representative quantitative survey of firm managers. But while their survey focused primarily on the price-setting decisions and cost structure within U.S. firms, we focus on the expectations of firm managers in New Zealand, as well as on their acquisition of information and their knowledge about monetary policy. This paper also builds on the work of Coibion, Gorodnichenko, and Kumar (2015),

which first utilized the New Zealand survey of firms. We differ from the latter primarily in that we focus on previously unexplored survey questions bearing on managers' knowledge of monetary policy. We also make use of a new wave of the survey that extracts, among other new questions, narrative answers from managers about the formation of their expectations.

Finally, our work relates closely to the literature on the nature of the formation of expectations among different economic agents. This literature explores the need for moving beyond the assumption of full-information rational expectations. Robert Lucas (1972), Greg Mankiw and Ricardo Reis (2002), Michael Woodford (2003), Christopher Sims (2003), Gorodnichenko (2008), Bartosz Mackowiack and Mirko Wiederholt (2009), and Fernando Alvarez, Francesco Lippi, and Luigi Paciello (2011) all provide models that explore the implications of different ways of incorporating deviations from full-information rational expectations. Empirically, such deviations have already found repeated support. Mankiw, Reis, and Justin Wolfers (2004), for example, emphasize the time variation in disagreement across agents, while Coibion and Gorodnichenko (2012, 2015a) emphasize the predictability in different agents' forecast errors. Philippe Andrade and Hervé Le Bihan (2013) provide evidence that European forecasters update their forecasts infrequently and in a manner consistent with imperfect-information models. Christopher Carroll (2003) argues that information diffuses only gradually from professional forecasters to households. The direct evidence in this paper on how little firm managers know about the weights applied to different categories of goods in measuring aggregate price levels or about the inflation target of the RBNZ confirms that, at least for these agents, full information is also likely a poor approximation.

This paper is organized as follows. In section I we describe the survey. In section II we present evidence on the degree to which managers' expectations in New Zealand are anchored. In section III we study how much firm managers in New Zealand know about monetary policy. In section IV we focus on the sources of the differences in managers' inflation perceptions and forecasts, and in section V we extend those results to the United States. Section VI concludes.

I. Description of the Survey

We implemented a quantitative survey of firm managers' expectations about macroeconomic conditions in New Zealand, a comprehensive description of which may be found in Coibion, Gorodnichenko, and Kumar (2015). In this paper we discuss only the key features of the survey.

We executed the survey in five waves between September 2013 and August 2015. The first and largest wave was conducted between September 2013 and January 2014. After contacting around 15,000 firms, we attained a response rate of around 20 percent, and then recruited a sample of 3,153 firms to participate. Subsequent waves were accomplished by recontacting firms from the first wave. The second wave, implemented between February and April 2014, received 714 responses. The third and fourth waves, conducted in August–September 2014 and December 2014–January 2015, received 1,607 and 1,257 responses, respectively. In August 2015, we implemented a much smaller fifth wave (50 firms), aiming to collect responses to open-ended questions.

The main survey (first wave) focused on collecting a wide range of information on the characteristics of the firms, their price-setting decisions, and their expectations about recent and future inflation. Follow-up waves included some repeat questions from the main survey as well as new questions. The second wave, for example, not only asked firms about their beliefs of inflation but also about other macroeconomic indicators such as real GDP growth, unemployment, and interest rates. The third wave focused on collecting individual characteristics of the respondents, and the fourth wave explored how managers acquire and process new information. The smaller, fifth wave explored views on central bank credibility and knowledge about inflation and asked respondents to give narrative answers. With the exception of a handful of cases, the survey collected responses across waves from the *same* person within a given firm.

Firms were randomly chosen from four broad industries: manufacturing, retail and wholesale trade, construction and transportation, and professional and business services.[1] Phone interviews were done with the

1. The firm names and their basic details were purchased from the Kompass New Zealand and Knowledge Management Services databases. Prior to acquiring these, we used Statistics New Zealand data for 2012 to compute the proportion of firms that fall into each employment size group (6 to 19 workers, 20 to 49 workers, and more than 50 workers) for each sector so that we could match our population with the population of firms in the economy. For example, in the manufacturing industry in 2012, around 67 percent of firms had 6 to 19 workers, 21 percent had 20 to 49 workers, and 12 percent had more than 50 workers. Our survey population in the manufacturing industry contained similar proportions. For other industries, their respective employment size proportions were computed and our population was constructed accordingly.

general managers[2] approximately 10 days after the questionnaires had been emailed to them. Since manufacturing and professional and business services account for relatively large shares of GDP (according to 2012 Statistics New Zealand data), we aimed to have two-thirds of our sample from these two industries. The remaining third is a combination of firms from other industries. We excluded from the sample industries related to the government, community service, agriculture, fishing, mining, energy, gas, and water, since objectives and constraints of firms in these industries may differ from those of other firms due to differences in governance, regulation, and so on. The combined employment of firms in our sample represents about 5 percent of total employment in New Zealand. While our sample is not drawn to be perfectly representative of the New Zealand economy, we can use sampling weights to adjust for the size distribution of firms and the industrial composition.[3] For the smaller, fifth wave, we first grouped firms into four bins based on their inflation forecasts (0 to 3 percent, 4 to 6 percent, 7 to 10 percent, and more than 10 percent) in the previous waves. We then randomly selected firms from each group. Each group includes a similar number of firms and they are broadly representative of industry and size composition of firms in the overall sample.

The survey is unique both in its breadth of coverage and in the quantitative nature of the questions asked. While many surveys of firms exist, most tend to ask only qualitative questions. The few that ask quantitative questions tend to survey only a small and very unrepresentative group. For example, the RBNZ conducts a quarterly survey (namely, its Survey of Expectations) of a sample of economists and business and industry leaders. That survey started in the late 1980s, and its objective was to

2. Our objective was to contact top-level leaders at each firm. Of all respondents in our surveys, 97 percent are classified as managers, a category that includes managers, managing directors, directors, chief executive officers, and principal legal executives. The other respondents (3 percent) are marketing officers, industrial relations executives, product development officers, and so on.

3. See Coibion, Gorodnichenko and Kumar (2015) for more details, such as response rates, for each wave of the survey. Using sample weights makes little difference for our results. See table 1 and online appendix table A1. The distribution of firms by size across industries in the sample and in the population is given in online appendix tables A2 and A3. Online appendixes for papers in this volume may be found at the Brookings Papers web page, www.brookings.edu/bpea, under "Past Editions."

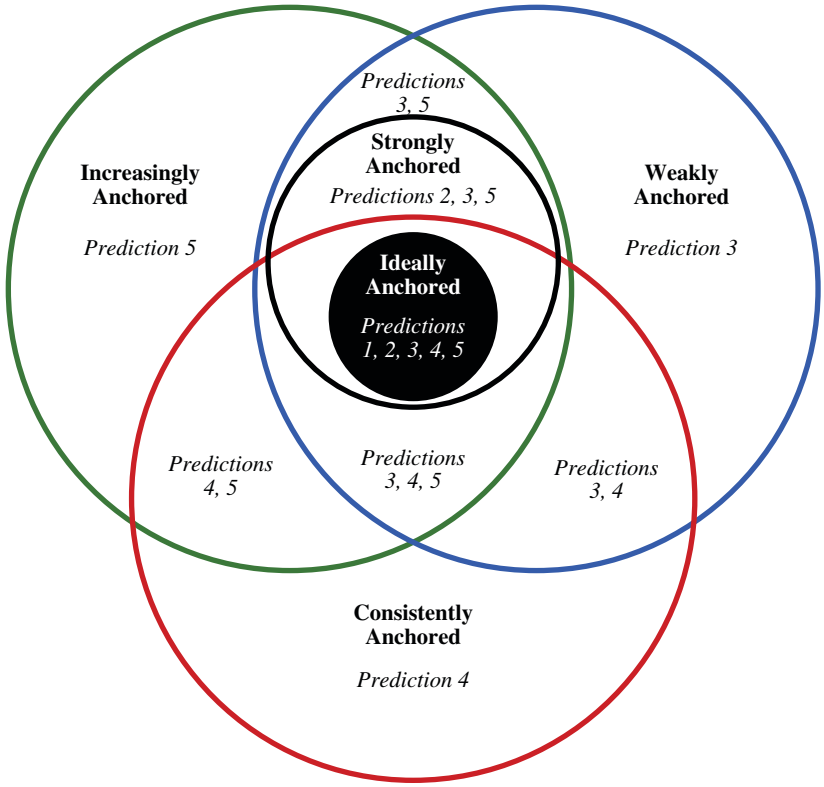
establish a database that might be useful for policy and research. Lars Svensson (2015) notes that inflation expectations reported in the RBNZ survey have been only slightly above actual levels of inflation for much of the inflation targeting period. However, the survey suffers from several issues. First, its sample size is very small and targets respondents mainly from the financial and business services industry.[4] Second, it is not purely at the firm level, since it also includes professional economists, market analysts, and public commentators, albeit those individuals make up a low proportion of the total. Third, the firms involved are typically very large ones, which is not at all representative of the New Zealand economy. For example, while firms with more than 100 employees represent less than 1 percent of all firms in New Zealand, they account for nearly all participants in the Survey of Expectations. Lastly, the sample is not random but, instead, is largely convenience-based.

II. Are Managers' Inflation Expectations Anchored in New Zealand?

Because there is no widely agreed-upon definition of "anchored" expectations, we consider five characteristics of inflation expectations that one expects to observe depending on the specific definition or extent of anchoring of expectations. As we show below, these predictions allow us to test five distinct definitions of anchored expectations, some of which are stronger than others. The first characteristic is average beliefs being close to the inflation target of the central bank. The second is beliefs not being too dispersed across agents. The third characteristic is that agents should be fairly confident in their forecasts and display little uncertainty, especially over the long run. Fourth, revisions in forecasts should tend to be small, especially at longer horizons. Fifth and finally, there should be little co-movement between long-run inflation expectations (which should be pinned down by the inflation target) and short-run inflation expectations (which should move with transitory shocks). In this section we assess each of these five predictions in turn and relate them to specific definitions of anchored inflation expectations. Figure 1 illustrates how the definitions are related to each of the five predicted characteristics in Venn diagram form.

4. The breakdown of the sample in the last quarter (2015Q2) is as follows: 91 respondents from financial and business services; 11 respondents from agriculture; 4 respondents from labor; and 8 respondents from other industries.

Figure 1. Definitions of Anchored Expectations and Associated Predictions[a]



a. This diagram shows how the five definitions of anchored expectations, and their predictions, relate to one another. Causality is established in the sense that if expectations are ideally ε_τ-anchored for all $\tau \geq 0$ given an arbitrary sequence of $\{\varepsilon_\tau\}_{\tau=0}^{\infty}$, then expectations are also strongly, weakly, consistently, and increasingly anchored according to $\{\varepsilon_\tau\}_{\tau=0}^{\infty}$, up to a scale.

For concreteness, suppose we have a measure of agents indexed by $i \in [0,1]$. Let π_t denote inflation in this economy at time t. Let $F^i_{t+\tau|t}(.)$ be the cumulative distribution function of i's time t belief about inflation at horizon $\tau \in \{0, \pm1, \pm2, \ldots\}$. Finally, let π^* be the central bank's inflation target. In this model, we only focus on bounded domain distributions for individuals, first because it is a reasonable assumption that no one assigns positive probability to infinite inflation or deflation, and second because it makes the analysis more intuitive. Nevertheless, the results can be extended to unbounded distributions with proper adjustments of definitions. Proofs for all lemmas and propositions that relate the definitions with predictions may be found in the online appendix.

II.A. Ideally Anchored Expectations

Definition 1. We say inflation expectations at time t for any horizon $\tau \geq 0$ are *ideally* ε*-anchored* if the support of every agent's belief of inflation at that time and horizon lies within ε of the central bank's target, meaning that

$$F^i_{t+\tau|t}(\pi^* + \varepsilon) - F^i_{t+\tau|t}(\pi^* - \varepsilon) = 1, \forall i \in [0,1].$$

Prediction 1: Average beliefs should be close to the inflation target. If inflation expectations at time t for any horizon $\tau \geq 0$ are ideally ε-anchored, then the average belief about inflation should lie within ε of the central bank's target, so that

$$bias_{t+\tau|t} \equiv \left| \overline{\pi_{t+\tau|t}} - \pi^* \right| < \varepsilon,$$

where $\overline{\pi_{t+\tau|t}} = \int_0^1 \pi^i_{t+\tau|t} di$ is the average belief across agents and $\pi^i_{t+\tau|t} \equiv E^i_t\{\pi_{t+\tau}\} = \int_{\mathbb{R}} x dF^i_{t+\tau|t}(x)$ is agent i's time t expectations of inflation at horizon τ.

Probably the most common interpretation of anchored expectations is that the average inflation forecast across agents, especially at longer horizons, remains stable and close to the central bank's inflation target (Ball and Mazumder 2014). Table 1 reports the average forecasts at the 12-month and the 5-to-10-year horizons for New Zealand firms in each survey for which these forecasts are available, as well as average managers' beliefs about inflation over the preceding 12 months. For comparison, we also report forecasts from the RBNZ, professional forecasts from Consensus Economics, and household forecasts from the Survey of Households produced by the RBNZ.

For each forecast horizon, firm managers' forecasts significantly exceed the inflation target as well as the forecasts of all other agents. They even exceed the forecasts of households, at least at 12-month horizons although not at long horizons.[5] Nor does the average manager forecast

5. The RBNZ first asks households if they understand what inflation means. Only those households who do (approximately half) are then asked to provide inflation forecasts. In the firm survey, all firm managers are asked to provide inflation forecasts. However, as documented in section IV, most managers can correctly explain the meaning of inflation.

Table 1. Inflation Forecasts of Firm Managers and Others for New Zealand, 2013Q4–2014Q4[a]

		Forecasts[b]							*Nowcasts/Backcasts*[d]			
			Professional forecasters		*Households*		*Firms*		*Households*		*Firms*	
Survey date	*Recent data*[c] *(1)*	*Central bank (2)*	*Mean (3)*	*SD (4)*	*Mean (5)*	*SD (6)*	*Mean (7)*	*SD (8)*	*Mean (9)*	*SD (10)*	*Mean (11)*	*SD (12)*
1-year inflation												
2013Q4	1.5	1.3	2.0	0.2	3.4	2.0	5.3	3.1	2.9	1.8	4.8	3.5
2014Q1	1.5	1.9	2.0	0.3	3.6	1.8	5.9	2.8	2.9	1.6	5.3	3.3
2014Q3	1.6	1.6	1.9	0.2	3.5	2.2	4.3	2.5	2.9	1.9	n.a.	n.a.
2014Q4	1.0	1.1	1.7	0.3	3.1	1.9	4.7	2.8	2.8	1.8	4.3	2.5
5-to-10-year inflation												
2014Q3	2.1	2.1	2.1	n.a.	4.1	2.8	3.4	2.4	—	—	—	—

Source: Consensus Economics, and the RBNZ's Survey of Households and Monetary Policy Reports.

a. Consistent with how the RBNZ calculates moments of household expectations, we exclude responses of firms' managers that are greater than 15 percent or less than −2 percent. Results for nontruncated data are reported in the online appendix table A1. Moments for firms are calculated without weights. Results with sample weights are also reported in the online appendix table A1. Standard deviations are expressed as percentage points, and all other values are expressed as percents.

b. Reports inflation forecasts of the RBNZ (column 2), mean forecasts from surveys, and the cross-sectional standard deviations of these forecasts (columns 3–8).

c. Reports most recently available inflation data at the time of the surveys.

d. Reports nowcasts/backcasts of inflation over the last 12 months.

appear particularly stable over time. Over the course of 2014, the average 1-year-ahead forecast of inflation fell by anywhere between 0.6 percentage point and 2.2 percentage points, depending on how one truncates the sample.[6] In contrast, professional forecasters, on average, reduced their forecasts by 0.3 percentage point over the same period, and households reduced them by 0.4 percentage point. While one must bear in mind the short time frame, this survey evidence suggests that the average forecast across firms is neither close to the inflation target nor stable over time.

It is possible that managers' average inflation beliefs were temporarily driven higher in New Zealand at the time of the survey by transitory changes in economic conditions that disproportionately affected their expectations relative to those of central bankers and those of households. For example, a boom in commodity prices could have raised inflation expectations temporarily. A weaker definition of anchored expectations could allow for time variation in average beliefs, but restrict the cross-sectional distribution of those beliefs.

II.B. Strongly Anchored Expectations

Definition 2. We say inflation expectations at time t for any horizon $\tau \geq 0$ are *strongly* ε-*anchored* if the support of every agent's inflation expectations at that time and horizon lies within ε of the average belief, so that,

$$F^i_{t+\tau|t}\left(\overline{\pi_{t+\tau|t}} + \varepsilon\right) - F^i_{t+\tau|t}\left(\overline{\pi_{t+\tau|t}} - \varepsilon\right) = 1, \forall i \in [0,1].$$

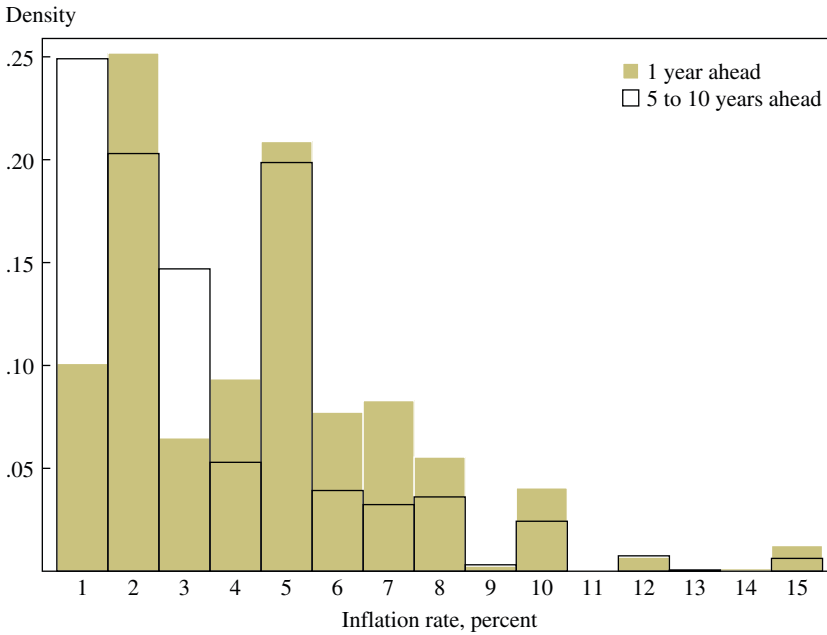
Prediction 2: Beliefs should not be too dispersed across agents. If inflation expectations at time t for horizon τ are strongly ε-anchored, then the dispersion of agents' beliefs about inflation should be less than ε, meaning that

$$sd_{t+\tau|t} \equiv \left[\int_0^1 \left(\pi^i_{t+\tau|t} - \overline{\pi_{t+\tau|t}}\right)^2 di\right]^{\frac{1}{2}} < \varepsilon.$$

Lemma 1. If inflation expectations are ideally $\varepsilon/2$-anchored, then they are strongly ε-anchored.

6. The RBNZ's survey of households drops all forecasts above 15 percent and below −2 percent. Our "truncated" sample applies the same restrictions, whereas our full sample includes all forecasts.

Figure 2. Distributions of Inflation Forecasts[a]



Source: Authors' survey of New Zealand firms.

a. Reports the distribution of 1-year-ahead and 5-to-10-years-ahead inflation forecasts in the third wave of the survey (2014Q3). The survey questions are: "During the *next twelve* months, by how much do you think overall prices in the economy will change? Please provide an answer in percentage terms." and "Over the next five to ten years, at what average percentage rate *per year* do you think that overall prices in the economy will be changing?"

Note that, as established by lemma 1, this is a weaker definition of anchored expectations. Table 1 reports the cross-sectional standard deviation in inflation forecasts at the same horizons as before. The dispersion in managers' 12-month-ahead inflation forecasts exceeds the dispersion among households every quarter, which is already an order of magnitude larger than that of professional forecasters. Hence, there is little concentration of managers' beliefs about one-year inflation forecasts. The distribution of forecasts for inflation at the 5-to-10-years horizon is illustrated in figure 2. While approximately 20 percent of managers forecasted a 2 percent rate, consistent with the RBNZ's long-run target, another 20 percent picked a much higher long-run forecast of 5 percent, and another 20 percent predicted that inflation over the next 5 to 10 years would average between 5 and 10 percent. This dispersion in beliefs about long-run inflation is particularly difficult to reconcile with anchored inflation expectations.

We now consider three additional predicted characteristics and associated definitions of anchored expectations, each of which is conceptually distinct and weaker than strongly ε-anchored expectations.

II.C. Weakly Anchored Expectations

Definition 3. We say inflation expectations at time t for any horizon $\tau \geq 0$ are *weakly* ε*-anchored* if the support of every agent's inflation expectations at that time and horizon lies within ε of that agent's belief, meaning that,

$$F^i_{t+\tau|t}\left(\pi^i_{t+\tau|t} + \varepsilon\right) - F^i_{t+\tau|t}\left(\pi^i_{t+\tau|t} - \varepsilon\right) = 1, \forall i \in [0,1].$$

Moreover, we say agent i is ε*-confident* of her forecast/backcast if her own perception of the degree to which her inflation expectation has varied is less than ε^2, that is:

$$E^i_t\left\{\left(\pi_{t+\tau} - \pi^i_{t+\tau|t}\right)^2\right\} < \varepsilon^2.$$

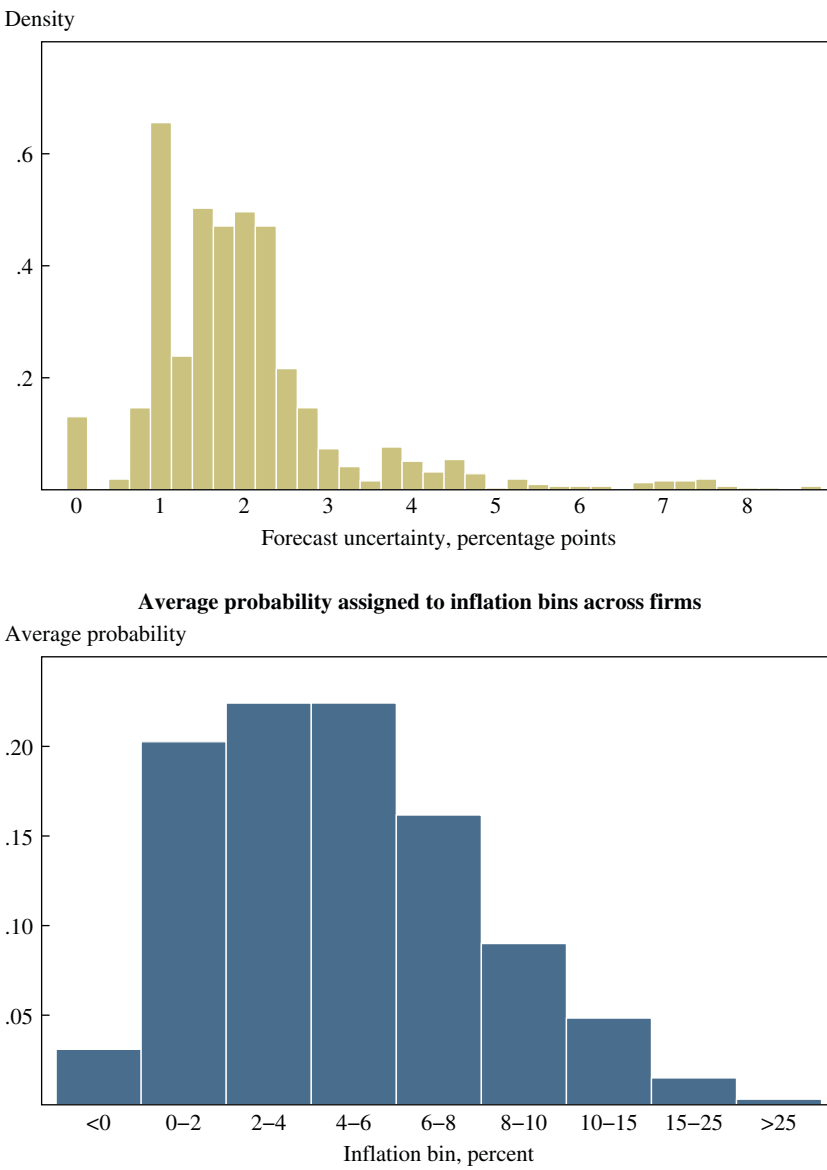
Prediction 3: Agents should show confidence in their forecasts. If inflation expectations are weakly ε-anchored for a given time and horizon, then all agents are ε-confident of their forecasts.

Lemma 2. If inflation expectations are ideally or strongly ε/2-anchored, then they are also weakly ε-anchored.

Under weakly anchored expectations, we now make no restrictions about agents having similar beliefs about long-run inflation or the central bank's target. Instead, the notion of targeting is now that each agent should be confident that inflation will be stabilized around whatever value each agent believes is the target. That is, one should perceive little risk of either high or low inflation in the future, so that the range of possible outcomes for inflation considered realistic by agents should be quite limited. Lemma 2 shows that this notion of anchoring is weaker than the previous two definitions.

In the fourth wave of the survey, we asked managers to assign probabilities to a wide range of possible inflation outcomes, allowing us to characterize the degree of uncertainty in their forecasts. From these distributional answers, we computed the standard deviation of each manager's forecast, and we report the distribution of these standard deviations in the upper panel of figure 3. The average standard deviation is 2 percentage points, so the firm managers on average report a lot of uncertainty around their

Figure 3. Uncertainty in Managers' Inflation Forecasts[a]



Source: Authors' survey of New Zealand firms.

a. The survey question asks managers to assign probabilities to nine bins. For each respondent, we construct a measure of uncertainty as the standard deviation of the reported distribution.

forecasts. As the upper panel also makes clear, there is considerable heterogeneity in the degree of uncertainty associated with individuals' forecasts. The lower panel reports the average probability assigned by managers to each bin of the distribution. While most of the mass is assigned to bins ranging from 0 to 6 percent inflation, much higher inflation rates receive a significant weight on average as well.

An alternative definition of anchored expectations, which is again weaker than strongly ε-anchored, is to allow agents to have different beliefs about inflation targets. In this case one can observe both a mean forecast that deviates from the true target as well as a high dispersion in forecasts (since agents disagree about targets). This is formalized in definition 4, which follows.

II.D. Consistently Anchored Expectations

Definition 4. We say inflation expectations of agent i for any horizon $\tau \geq 0$ are *consistently* ε*-anchored* at *t,* if the total change in the cumulative distribution function of his belief from $t-1$ to t is less than ε in magnitude:

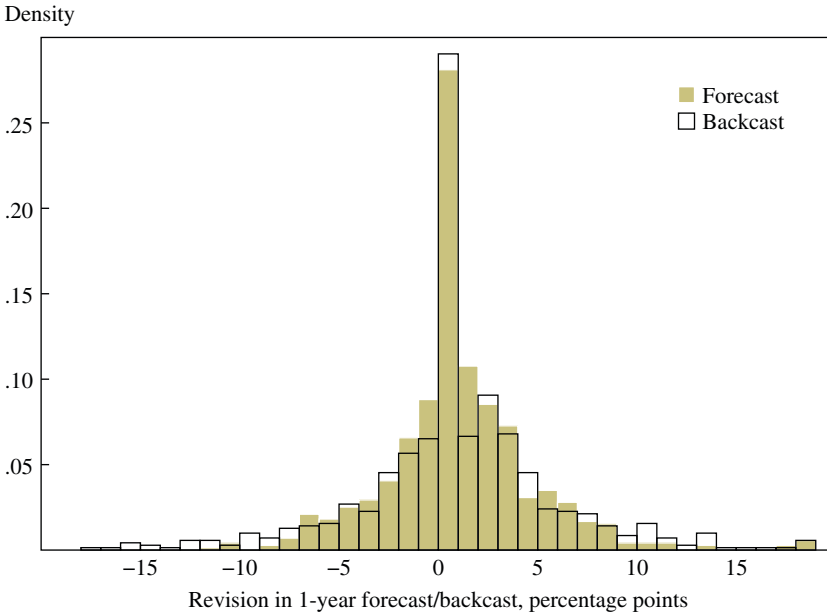
$$\int_{\mathbb{R}} \left| F^i_{t+\tau|t}(x) - F^i_{t+\tau-1|t-1}(x) \right| dx < \varepsilon.$$

Prediction 4: Agents should display small forecast revisions. The size of an agent's forecast revision of inflation at time t for any horizon $\tau \geq 0$ is less than ε if her inflation expectation for horizon τ is consistently ε-anchored at *t.*

Lemma 3. Define forecast revision for agent i at time t for horizon τ as $FR^i_{t+\tau|t} = \pi^i_{t+\tau|t} - \pi^i_{t+\tau-1|t-1}$. If inflation expectations for horizon τ are ideally ε/2-anchored at $t-1$ and t, then they are also consistently ε-anchored. Moreover, if expectations for horizon τ are strongly ε/2-anchored at $t-1$ and t, then they are also consistently $(\varepsilon + \delta')$-anchored where $\delta' \equiv \left| \int_0^1 FR^i_{t+\tau|t} di \right|$ is the absolute size of average forecast revision across agents.

Under this definition of anchoring, individuals' revisions in their inflation forecasts should tend to be small, as established in prediction 4, since agents expect the central bank to be able to keep inflation stable over long enough horizons. Because our survey includes a panel dimension, we can examine this prediction using revisions in managers' inflation forecasts. In figure 4, we plot the distribution of revisions of managers' 1-year inflation forecasts along with (for comparison) the distribution of revisions in their views about inflation over the previous 12 months. While one might expect the latter to display significantly more dispersion, the figure illustrates that the dispersion in revisions of forecasts at the 1-year horizon is almost as

Figure 4. Revisions of Managers' One-Year Inflation Forecasts and Backcasts



Source: Authors' survey of New Zealand firms.

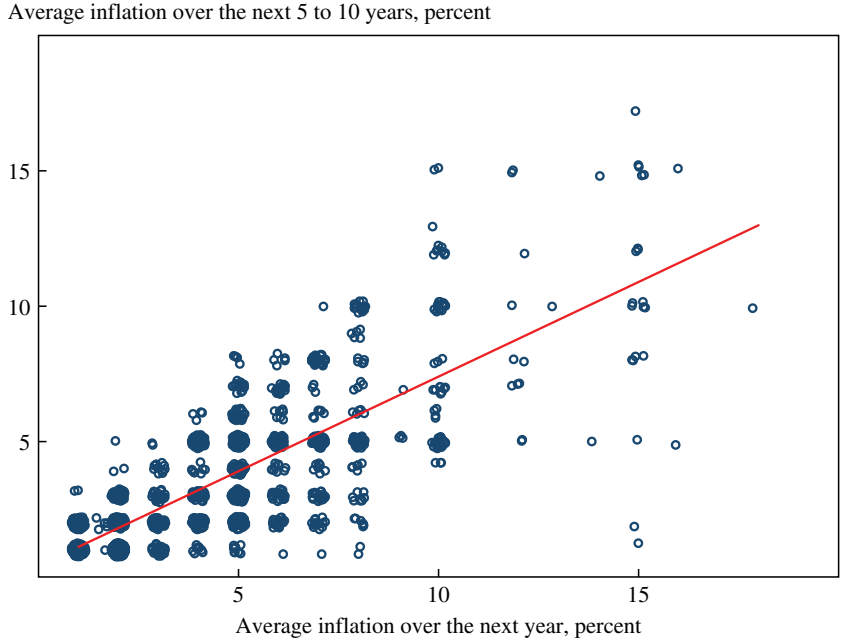
large: it is common in the survey to see revisions in 1-year forecasts of inflation that are 5 percentage points or more in absolute value. Given the stability of inflation in New Zealand over this period, it is difficult to reconcile such dramatic revisions with any notion—weak or strong—of well-anchored expectations.

II.E. Increasingly Anchored Expectations

A final interpretation of anchored expectations comes from looking at the correlation of short-run and long-run expectations. Consider an extreme example of anchoring: if central banks are able to successfully stabilize beliefs about long-run inflation to a target, then short-run and transitory fluctuations in inflation and short-run inflation expectations should be unrelated to these longer-run beliefs. This is established more formally in the following definition and prediction.

Definition 5. Given a sequence $\{\varepsilon_\tau\}_{\tau=0}^{\infty}$ at time *t,* we say inflation expectations are *increasingly T-anchored* at time *t* if for any $\tau \geq T$, expectations are strongly ε_τ-anchored.

Figure 5. Short-Term vs. Long-Term Inflation Expectations[a]



Source: Authors' survey of New Zealand firms.

a. Shows the relationship between 1-year-ahead and 5-to-10-years-ahead inflation forecasts in the third wave of the survey (2014Q3). The survey questions are: "During the *next twelve* months, by how much do you think overall prices in the economy will change? Please provide an answer in percentage terms." and "Over the next five to ten years, at what average percentage rate *per year* do you think that overall prices in the economy will be changing?"

Notice that increasingly *T*-anchored expectations are weaker than strongly anchored expectations when they are required to be strongly ε_τ-anchored for $\tau \geq 0$, since this condition does not impose any restrictions on expectations over short-run horizons.

Prediction 5: Long-run expectations should be unpredictable using short-run expectations. Consider the following regression for time *t* data:

$$\pi^i_{t+\tau|t} = \alpha_\tau + \beta_\tau \pi^i_{t+1|t} + error_t.$$

Now, given the sequence $\{\varepsilon_\tau\}_{\tau=0}^{\infty}$ such that $\lim_{\tau\to\infty}\varepsilon_\tau = 0$, suppose expectations are increasingly *T*-anchored for an arbitrary $T \geq 1$. Then $\lim_{\tau\to\infty} \beta_\tau = 0$.

Figure 5 documents that this prediction does not hold in our data: Managers who expect higher short-run inflation also tend to expect higher

long-run inflation. While one might still expect to find a positive slope between the two sets of expectations simply because the time horizon of the long-run expectations is not long enough (τ is too small), the size of the slope coefficient that we find is too large to be explained by this feature of the data. For example, if agents set their 5-to-10-years-ahead forecasts equal to a weighted average of their 1-year-ahead expectation and their expectation about long-run inflation, the implied slope of the relationship should be small (between 0.1 and 0.2 for the 5-to-10-years-ahead forecasts), whereas figure 5 illustrates a slope of 0.70. Consequently, the strength of the relationship between managers' long-run and short-run inflation expectations is much greater than one would expect to see under well-anchored expectations.

In short, the survey of firm managers in New Zealand suggests that no matter which definition of anchoring we adopt, and even using much weaker versions than commonly considered, the inflation expectations of firm managers do not conform in the least bit to the properties one would expect from agents with well-anchored expectations.

III. Credibility vs. Knowledge of the Central Bank's Objectives

That inflation expectations in New Zealand appear to be unanchored despite 25 years of inflation targeting and relatively stable inflation is puzzling. One reason so many managers' long-run inflation forecasts are so high might be that they do not view the RBNZ as a credible institution, that is, they do not believe in the RBNZ's ability or willingness to achieve its long-run inflation objective. Another possibility is that many managers are unaware of the objectives of the central bank and of recent inflation dynamics. In this section, we try to differentiate between these two potential explanations.

III.A. Managers' Knowledge

In the fourth and fifth waves of the survey, we asked managers several questions designed to assess their knowledge of the objectives of the central bank and of monetary policy more generally. First, we posed the following question to them in the fourth wave of the survey:

What is the main objective of the Reserve Bank?

a. Keep the exchange rate stable	[*23 percent*]
b. Promote full employment	[*25 percent*]
c. Keep interest rates low and stable	[*11 percent*]
d. Keep inflation low and stable	[*31 percent*]
e. Help the government finance its spending	[*10 percent*]

The percent figures in square brackets show the shares of responses.[7] Only 31 percent of respondents correctly chose answer (d). Given that there were five choices available, this suggests that very few people know even in a broad sense the main objective of the RBNZ. We then asked the following question:

> What annual percentage rate of change in overall prices do you think the Reserve Bank of New Zealand is trying to achieve?
> *Answer:* . . . . . percent

We present a detailed distribution of the responses to that question in table 2. Of the respondents, only 12 percent correctly responded 2 percent, although an additional 25 percent said either 1 percent or 3 percent, the bottom and top of the target range of the RBNZ. At the same time, 15 percent of respondents said the RBNZ's target inflation rate was 5 percent and 36 percent of them said it was higher than 5 percent, including 5 percent of respondents who said it was 10 percent or more.

Finally, we asked managers the following:[8]

> What is the name of the Governor of the Reserve Bank of New Zealand?
> a. Graeme Wheeler [*30 percent*]
> b. Alan Bollard [*39 percent*]
> c. Bill English [*17 percent*]
> d. Charles Cowley [*9 percent*]
> e. I don't know [*4 percent*]

In that list, only two people (Wheeler and Bollard) are or were affiliated with the RBNZ, and they obtained 69 percent of the responses. The correct answer (a) was again chosen by only 30 percent of respondents. The most popular response was the name of the governor (Bollard) who had actually stepped down from the office more than two years before the survey.

Jointly, these questions point toward a pervasive lack of knowledge about monetary policy on the part of firm managers. In fact, only 10 percent of respondents answered all three questions exactly correctly. But those who answered all questions correctly displayed significantly

7. Because of the large sample size, both here and in the remainder of the paper we can reject the null hypothesis that the share of managers picking a given option is equal to $1/N$, where N is the number of options in a multiple choice question.

8. Graeme Wheeler was the governor of the RBNZ at the time of the survey (2015Q1). He had been the governor since September 2012. Alan Bollard had been the governor of the RBNZ until September 2012. Bill English was the minister of finance at the time of the survey. Charles Cowley is a randomly chosen person in New Zealand.

Table 2. Perceived RBNZ Inflation Target and Managers' Inflation Forecasts

		Inflation forecast							
		Mean[a]			*Standard deviation*[b]			*Average uncertainty*[c]	
Perceived inflation target of the RBNZ	*Percent of managers reporting a target* (1)	*1 year, wave 4* (2)	*1 year, wave 3* (3)	*5–10 years, wave 3* (4)	*1 year, wave 4* (5)	*1 year, wave 3* (6)	*5–10 years, wave 3* (7)	*1 year, wave 4* (8)	*Average inflation backcast*[d] (9)
1	14.96	3.1	2.7	2.2	2.2	2.0	1.8	1.64	2.6
2	12.41	2.8	2.5	2.1	1.5	1.4	1.3	1.68	2.3
3	10.26	4.1	3.6	3.0	2.4	2.1	1.9	1.91	3.1
4	11.46	5.7	5.2	4.1	3.2	2.9	2.7	2.24	5.6
5	14.96	5.9	5.3	4.1	3.2	2.7	2.4	2.10	6.1
6	8.91	5.4	5.0	3.9	2.9	2.7	2.8	2.05	6.1
7	7.80	5.8	5.4	3.7	3.6	2.8	2.2	2.16	6.3
8	8.75	6.7	6.2	4.8	3.6	3.0	3.2	2.38	6.6
9	5.49	5.7	5.2	4.0	3.1	2.5	2.6	2.02	7.3
10+	5.01	6.4	6.0	5.0	2.9	2.8	3.0	2.05	8.2

Source: Authors' survey of New Zealand firms.

a. Reports mean inflation forecasts at different horizons for various waves of the survey.

b. Reports cross-sectional standard deviation of inflation forecasts at different horizons for various waves of the survey.

c. Reports average (across firms) standard deviation of inflation forecast implied by the probability distribution reported by firms' managers.

d. Reports the average perceived inflation rate over the previous 12 months.

better-behaved forecasts (in the sense of being closer to those of professionals): Their mean long-run inflation forecast was exactly 2 percent with a cross-sectional standard deviation of just 1.2 percentage points, whereas all other respondents had a mean forecast of 3.7 percent with a standard deviation of 2.6 percentage points. Likewise, the mean short-term forecast of these all-correct-answer respondents was 2.6 percent (standard deviation of 1.4 percentage points), while all other respondents had a mean of 5.2 percent (standard deviation of 3.2 percentage points). These respondents also had much better knowledge of recent inflation dynamics, with average backcasts of just 1.5 percent (standard deviation of 0.6 percentage point) compared with 4.6 percent (standard deviation of 2.4 percentage points) for other respondents.

III.B. Managers' Views of the Central Bank's Credibility

Additional evidence does not suggest that the RBNZ suffers from a significant credibility problem. For example, among respondents who knew that the target inflation rate was centered at 2 percent, 89 percent forecasted that inflation over the next 5 to 10 years would range from 1 to 3 percent, as illustrated in table 2. Among those who knew that the main objective of the central bank was to keep inflation low and stable, the average long-run forecast of inflation was again 2 percent. This suggests that from the point of view of those agents who know the central bank's objective, that objective is credible, and it is embedded in their forecasts.

The credibility of the RBNZ extends even to many of those who were incorrect about its actual inflation target: managers who thought that the RBNZ's target was 3 percent reported an average long-run inflation forecast of 3 percent, and those who thought the target was 4 percent reported an average long-run inflation forecast of 4 percent.[9] So while those two clusters of respondents were incorrect about the value the RBNZ was trying to achieve, their forecasts were consistent with the notion that the RBNZ would be able to achieve what they perceived to be its objective.

In the fifth wave of the survey, we posed two additional questions meant to directly address the credibility of the central bank. One question was this:

> Do you think the central bank can control inflation over the next 5 to 10 years?

9. For higher RBNZ inflation targets believed by respondents, the relationship between the target and managers' forecasts becomes flatter, with forecasts no longer rising one-for-one with targets but still increasing.

The point was to assess whether managers believed the central bank could achieve its inflation target over a medium to long time horizon. Out of the 50 respondents in the fifth wave, 49 responded *yes.* This indicates broad support on the part of managers for the view that monetary policymakers can achieve their medium- to long-run policy objectives. We also asked respondents the following question:

> Do you think the central bank can control inflation in the next 12 months or so?

Strikingly, 47 of the 50 respondents answered *yes.* Thus, the vast majority of firm managers assign tremendous credibility to the central bank, since they believe it can control inflation even at short horizons.

We interpret these results as suggesting that the credibility of the RBNZ is well established in New Zealand. The issue appears not to be one of credibility but, instead, that many managers are relatively uninformed about the practical objectives and targets of the central bank. This lack of quantitative information is reflected in the forecasts they report. Indeed, as documented in table 2, managers who were uninformed about the RBNZ's target inflation rate also tended to be much less informed about recent inflation dynamics, and their forecasts were also associated with much more uncertainty. Given the ease with which information about monetary policy can be accessed, it may seem surprising that so many managers of firms are not more informed about it.

III.C. Accounting for the Lack of Knowledge

What could account for these differences in knowledge about monetary policy? One possibility could be that better-educated managers simply know much more about monetary policy than others, but as column 7 of table 3 reveals, there are few differences in the average education levels across groups. Another possibility is that different types of managers place different values on information about monetary policy or about inflation more generally. To assess this, we asked managers about their willingness to pay for monthly forecasts of inflation as well as for forecasts of other macroeconomic variables. The results, presented in columns 2 and 3 of table 3, indicate that on average, firm managers who correctly reported low values of the RBNZ's inflation target also reported being willing to pay much more for inflation forecasts, both in dollar terms and as a share of firm sales, than managers who reported higher target values. By contrast, no such striking pattern exists for other macroeconomic variables (columns 4 and 5). This finding confirms that managers do indeed seem to

Table 3. Value of News for Firm Managers in New Zealand

Perceived inflation target of the RBNZ	*Percent of managers reporting a target (1)*	*Willingness to pay for having access to professional forecast, by variable*[a]				*Absolute slope of the profit function*[b] *(6)*	*Average years of schooling (7)*	*Frequency of following media news about the economy*[c] *(8)*	*Frequency of reading public sources about the economy*[d] *(9)*
		Inflation		*GDP (dollars per year) (4)*	*Unemployment (dollars per year) (5)*				
		Dollars per year (2)	*Percent of sales (3)*						
1	14.96	195	0.018	125	118	1.08	15.74	3.7	6.7
2	12.41	202	0.017	130	121	1.13	15.61	4.2	7.3
3	10.26	172	0.015	126	126	1.13	15.22	6.9	10.1
4	11.46	126	0.010	149	129	1.01	15.63	9.8	12.9
5	14.96	110	0.007	151	127	1.01	15.24	10.9	14.6
6	8.91	106	0.008	143	136	0.98	15.16	11.7	14.8
7	7.80	104	0.007	149	137	1.00	15.36	11.4	14.3
8	8.75	107	0.008	150	134	0.89	15.35	11.5	14.5
9	5.49	101	0.007	165	132	0.93	15.12	11.8	14.6
10+	5.01	98	0.006	141	127	1.01	15.35	11.5	15.0

Source: Authors' survey of New Zealand firms.

a. Columns 2, 4, and 5 report the average dollar amount managers are willing to pay per year for monthly professional forecasts of economic variables. Column 3 reports the average percent of annual sales managers are willing to pay for monthly professional forecasts of inflation.

b. Reports the average absolute slope of the profit function with respect to the price of a firm's main product.

c. Reports the frequency (months) at which managers update their information about the state of the economy using media news.

d. Reports the frequency (months) at which managers check public sources about the state of the economy.

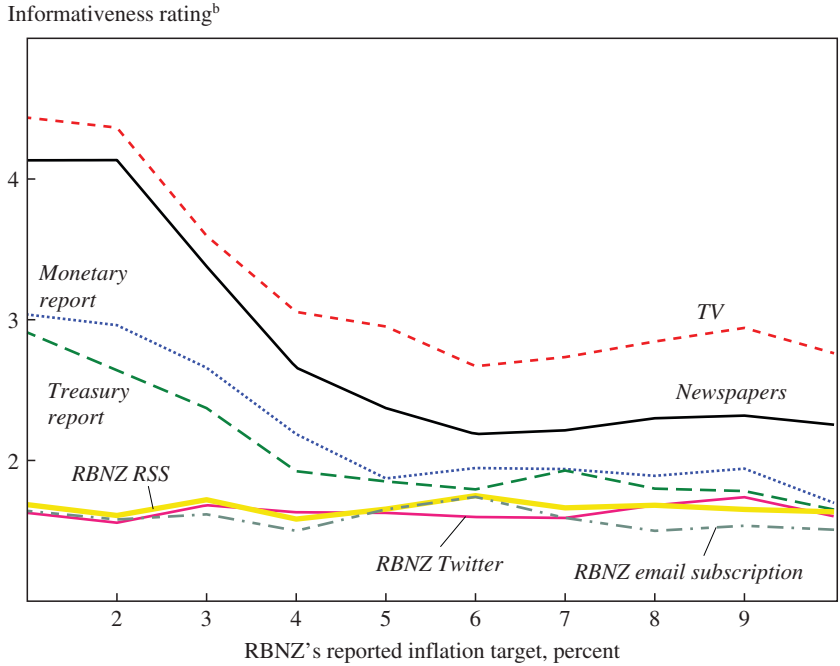
assign very different values to information about inflation, and that these valuations are reflected in their knowledge of both monetary policy and inflation dynamics.

Also consistent with an information channel is the frequency with which managers follow news about the economy. We asked managers about the frequency (daily, weekly, monthly, quarterly, semiannually, annually, less than annually) with which they followed news about the economy from media sources as well as, in a separate question, from official public sources (such as monetary policy reports). Our results, converted into answers in months, are presented in columns 8 and 9 of table 3 for each group of managers that gave a common answer as to the RBNZ's inflation target. Managers who were aware of the low inflation target followed media reports closely (once every 3 to 5 months on average) but this attention to news declines sharply with higher inflation targets, before stabilizing for managers who reported targets of 5 percent or more, for whom the average frequency in following media reports was approximately once per year. An identical pattern occurs with news from public sources, albeit at lower frequencies.

We also asked firm managers to describe how informative for their business decisions (on a scale of 1 for "irrelevant" to 6 for "extremely worthwhile") they found different types of news reports to be: television, newspapers, monetary policy reports, and different kinds of direct communications from the RBNZ. Results are plotted in figure 6, averaged across managers depending on their answers about the RBNZ's target rate of inflation. Managers who said the target was either 1 or 2 percent reported that television and newspapers were quite useful to their business decisions, consistent with their frequent use of these media as documented in table 3, and also reported that monetary and Treasury reports were somewhat useful. However, the perceived usefulness of all these media for business decisions declines sharply for managers who reported higher inflation targets. Meanwhile, none of the managers, regardless of their belief about the inflation target, reported finding much utility for their business decisions from direct communications by the central bank via email, RSS, or Twitter.

Why might different managers perceive the value of information about inflation to be different? One possibility is that the characteristics of the firms in which they work influence their perception of the value of information. For example, as column 6 of table 4 documents, there are pronounced differences in the average absolute slope of the firm's profit function with respect to firm's price, as defined in Coibion, Gorodnichenko, and Kumar (2015), depending on how managers responded to questions about the

Figure 6. Informativeness of Different Types of Media[a]



Source: Authors' survey of New Zealand firms.

a. Plots the mean informativeness for decisions of each type of media for all managers who reported that the inflation target for the RBNZ was one of the values on the x-axis.

b. Based on a survey rating scale of 1 to 6, with 1 being "irrelevant" and 6 being "extremely worthwhile."

RBNZ's inflation target.[10] Steeper profit functions imply that information should be more valuable to the firm and, consistent with this incentive effect, we find that managers who report low values of the RBNZ's target rate (and therefore have better information about monetary policy) also tend to work for firms whose profit functions are steeper and therefore where information is more valuable. This suggests that the characteristics of their own firms might be important in explaining the underlying differences in how managers value information about inflation and monetary policy.

We investigate whether the characteristics of managers and firms are correlated with a manager's knowledge of monetary policy more formally,

10. The slope of the profit function is calculated as follows: A firm is asked to report by how much (in percent) it would change the price of its main product if it were to do so for free and by how much this price change would translate into increased profits as a share of firm's revenue. The slope is the change in profit divided by the percent change in the price.

Table 4. Correlates of Knowledge of Monetary Policy[a]

	Absolute error in the perception of the central bank's target rate of inflation			*Count of wrong answers about the objective, target, and governor of the RBNZ*[b]		
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Firm characteristics						
Log (age)	0.086		0.052	0.077		0.129*
	(0.077)		(0.129)	(0.059)		(0.074)
Log (employment)	0.216**		0.259**	0.259***		0.210***
	(0.092)		(0.128)	(0.070)		(0.073)
Labor's share of costs	–0.010		–0.014	–0.010		–0.014**
	(0.007)		(0.010)	(0.006)		(0.006)
Foreign trade share	0.021***		0.018***	0.009***		0.008***
	(0.004)		(0.005)	(0.003)		(0.003)
Number of competitors	–0.011**		–0.017***	–0.006		–0.008**
	(0.005)		(0.006)	(0.004)		(0.004)
Average profit margin	0.013*		0.006	0.001		–0.000
	(0.007)		(0.010)	(0.006)		(0.006)
Price relative to competitors	0.011*		0.011	0.004		0.004
	(0.006)		(0.008)	(0.005)		(0.005)
Absolute slope of profit function	–0.226***		–0.142	–0.056		–0.081
	(0.087)		(0.138)	(0.075)		(0.080)

(*continued on next page*)

Table 4. Correlates of Knowledge of Monetary Policy[a] (*Continued*)

	Absolute error in the perception of the central bank's target rate of inflation			*Count of wrong answers about the objective, target, and governor of the RBNZ*[b]		
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Manager characteristics						
Age		−0.014*	−0.014		−0.008	−0.007
		(0.007)	(0.009)		(0.005)	(0.005)
Years of schooling		−0.069***	−0.051*		−0.042**	−0.041**
		(0.023)	(0.030)		(0.018)	(0.019)
Tenure		0.038***	0.001		0.026***	−0.006
		(0.008)	(0.014)		(0.006)	(0.008)
Income		0.003*	0.001		0.001	0.001
		(0.001)	(0.002)		(0.001)	(0.001)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	1,253	1,138	1,141	1,256	1,141	1,141
R^2	0.243	0.194	0.179	0.154	0.122	0.161

Source: Authors' survey of New Zealand firms.

a. Reports results of regressing errors about objectives, target, and governor of the RBNZ on a set of firm and manager characteristics indicated in the left column. Robust standard errors in parentheses. Statistical significance at the ***1 percent, **5 percent, and *10 percent levels.

b. Survey asked respondents about the RBNZ's objective, inflation target, and current governor (see subsection III.A). If all answers are correct, the count is equal to 0; if all answers are wrong, the count is equal to 3.

as follows. We consider two types of errors made by managers: errors specific to the target (the absolute deviation of their perception of the RBNZ target rate from 2 percent) and overall errors in answering questions (the sum of the number of answers they got wrong on the three survey questions on monetary policy). We then regress each of these manager-specific errors on a set of firm-specific characteristics, including the firm's age, employment at the firm, labor's share of total costs, foreign sales as a share of total revenues, the number of competitors reported by the manager, the firm's average profit margin, the price of the firm's main product relative to that of its competitors, and the absolute slope of the firm's profit function. All but the last variable come directly from survey questions asked of managers, and the slope is constructed from other questions asked of the managers. We also control for individual characteristics of the manager, such as their age, years of schooling, income, and tenure at the firm. Finally, we include industry fixed effects.

A few results (table 4) stand out as particularly robust. First, the number of competitors faced by a firm is systematically associated with smaller errors about the RBNZ's inflation target and about monetary policy more generally. One might interpret this as higher competition inducing managers to pay more attention to economic conditions, including monetary policy, to avoid being driven out of business. We also find that firms that receive more of their sales from abroad make bigger errors about monetary policy in New Zealand, which likely reflects their reduced incentive to track New Zealand's economy relative to that of their trading partners. Having more years of schooling on the part of managers is systematically associated with smaller errors about monetary policy. Interestingly, managers of larger firms make larger errors both about the RBNZ's inflation target as well as about monetary policy more broadly. Hence, these results confirm that characteristics of the firm play some role in accounting for the knowledge of monetary policy and inflation that managers display. However, because much of the variation remains unexplained, we turn to a more narrative approach to assess in greater depth how managers form and use their inflation expectations.

IV. What Managers Know about Inflation, How They Learn It, and Whether It Matters

Since the properties of firm managers' inflation expectations and perceptions appear to be so at odds with anchored expectations, we consider in this section three general sets of questions. First, what do managers actually

know about inflation? Perhaps many are confused about the concept or do not understand how it is measured. Second, where do managers get the information that goes into their beliefs about inflation and inflation expectations? Does it come from professional forecasters and statistical agencies, or from the media, or is it based on their daily experience with prices, either through their professional experience with competitors and clients or through their own shopping experience? Third, do their inflation expectations matter for any of the decisions that they make as managers? We address each of these in turn.

IV.A. What Do Managers Know?

Given the properties of firm managers' perceptions and expectations of inflation described in the previous sections, one might wonder to what extent they are knowledgeable about what inflation means or how it is measured as well as what sources of information they rely on to formulate their inflation expectations. The smaller, fifth wave of the survey was designed to address these points, using a combination of quantitative and narrative questions. Because narrative questions are much harder to implement on a vast scale, we restricted the sample to 50 firms drawn from different bins of the inflation expectation distribution (roughly 12 firms each from bins of firms that had previously forecasted inflation of 0–3 percent, 4–6 percent, 7–10 percent, and greater than 10 percent). As before, we asked managers of these firms to state their inflation expectations over the next 12 months and their perceptions of inflation over the previous 12 months.

We first assessed their basic knowledge of the term "inflation" by asking them "What is your understanding of the term inflation?" This is the same open-ended question that is posed to households in the RBNZ survey. We graded managers' narrative answers on a scale ranging from 0 to 2 points. Our reference answer was this: "Inflation is *the increase* (or *change*) [1 point] *in the general (average) price level of goods and services* [1 point] in the economy." Hence, managers received one point for recognizing that inflation captures a change in price levels and a second point for recognizing that it measures economy-wide prices.

Using this grading scheme, no managers received a grade of 0. Only eight managers received a score of 1, leaving 84 percent of managers in this sample as able to correctly define inflation. Of the eight who made a mistake, three stated that inflation measured food or "basic" commodity prices. The remaining five made statements inconsistent with inflation

capturing the change in prices.[11] There was little difference in the perceived inflation rates or expectations of managers across the two groups, suggesting that these differences in understanding of the concept of inflation play little role in accounting for the heterogeneity in managers' forecasts. In contrast, households in the RBNZ survey fail to correctly identify inflation at much higher rates. For example, in the May 2015 survey, only 52 percent of respondents could identify inflation. So firm managers display a much better understanding of the meaning of the term *inflation* than households.

Another reason why firm managers might hold inflation expectations and perceptions so at odds with recent inflation measurements is that they do not believe the official inflation statistics. To address this possibility, we asked firm managers in the fifth wave of the survey the following question:

> Do you think official inflation data are credible in the sense that it reflects the true rate at which overall prices in the economy change? [*Yes/No*]

Eighty-six percent of managers responded that official inflation data are credible, and 14 percent expressed skepticism. The average inflation forecast among managers who did not believe the official inflation data is only 1 percentage point higher than that among managers who did believe the official data. As a result, skepticism about the quality of official inflation statistics can explain neither the high mean of managers' inflation forecasts nor the dispersion in those forecasts.

If firm managers understand the concept of inflation and believe that government officials correctly measure it, how can they then perceive levels of inflation so different from what is measured by statistical agencies? Given that aggregate inflation is a weighted average for different categories of goods, two nonexclusive explanations are possible. One is that managers assign different weights to categories of goods from those assigned by statistical agencies (such as overweight gasoline price movements). Another is that managers are mistaken about the sizes of price changes for certain categories of goods (for example, they might think food prices went up more than they did). To assess these two explanations, we asked managers in the fifth wave to report how much weight they believe statistical agencies assign to different categories of goods and services when constructing

11. For example, one manager stated that inflation is the average price in the economy. The others said something akin to "when prices change, inflation changes."

overall price measures for New Zealand.[12] The specific set of categories includes housing prices, stock prices, food prices, health care costs, gasoline prices, the cost of rent, and car prices. We then asked managers to report their beliefs about price changes over the last 12 months for each of the same categories.

The results are presented in table 5, along with the actual weights applied to these categories in the construction of the Consumer Price Index (CPI) and the actual inflation rates for each category of goods. Managers' average beliefs about the weights on housing prices (22 percent, rounding to the nearest whole number), stock prices (8 percent), and gasoline prices (19 percent) far exceed the true values (4, 0, and 5 percent respectively). In contrast, managers significantly underestimated the weight assigned to food prices (believing it to be 9 percent whereas its true weight in the CPI is 19 percent). There are also some notable differences between managers' beliefs about inflation at the category levels and actual inflation rates. Managers significantly underestimated inflation in stock prices (believing it to be 4 percent whereas the actual change was 13 percent) but overestimated inflation in car prices (by 7 percentage points) and food prices (by 4 percentage points).

To quantify the relative importance of these two channels, it is useful to introduce some notation. We denote the actual weight for subcategory s with w_s^a and perceived weight for firm i with $w_{s,i}^p$. Likewise, we define the actual inflation rate for subcategory s with π_s^a and perceived inflation for firm i with $\pi_{s,i}^p$.

The contribution of price changes in these specific categories to actual aggregate inflation is

$$\tilde{\pi}^a = \sum\nolimits_s w_s^a \pi_s^a.$$

Equivalently, we construct firm i's perceived contribution of these categories to aggregate inflation as

$$\tilde{\pi}_i^p = \sum\nolimits_s w_{s,i}^p \pi_{s,i}^p.$$

12. The specific phrasing of the question was, "How much weight do you think statistical agencies place on each of the following categories of prices when constructing overall price measures for New Zealand? (these do *not* need to sum to 100 percent). Please provide percentage answers."

Table 5. Managers' Perceptions of Construction of Inflation Index vs. Actual Construction[a]

	Perceived		*Actual*	
	Weight	*Inflation rate*	*Weight*	*Inflation rate*
Housing prices	21.7	7.5	4.4	5.3
	(14.3)	(1.6)		
Stock prices	8.4	3.7	0.0	13.3
	(5.7)	(2.8)		
Food prices	8.9	4.0	18.8	0.2
	(8.1)	(2.0)		
Health care costs	8.4	0.7	4.0	1.8
	(6.4)	(1.7)		
Gasoline prices	18.9	–6.3	4.7	–7.4
	(12.7)	(6.0)		
Cost of rent	9.4	3.5	9.4	2.3
	(6.4)	(1.8)		
Car price	10.3	3.3	3.3	–3.8
	(9.7)	(1.6)		
Implied inflation contribution[b]	1.79 (2.12)		0.08	
Error[c]	1.71 (2.12)			
Of which:				
$\sum_s w_s^a(\pi_{s,i}^p - \pi_s^a)$	1.18 (0.48)			
$\sum_s(w_{s,i}^p - w_s^a)\pi_s^a$	0.78 (1.48)			
$\sum_s(w_{s,i}^p - w_s^a)(\pi_{s,i}^p - \pi_s^a)$	–0.24 (1.85)			

Source: Statistics New Zealand and authors' survey of New Zealand firms.

a. The top panel of the table shows perceived and actual weights and inflation rates for the subcategories of the CPI in the left column. In all panels, standard deviations are in parentheses.

b. Shows the total perceived ($\tilde{\pi}^p$) and actual ($\tilde{\pi}^a$) percentage-point contribution of the subcategories in the left column to aggregate CPI inflation.

c. Shows the difference between total perceived and actual contribution to aggregate CPI inflation and its decomposition.

The error that firm *i* makes about this contribution can be decomposed as:

$$\tilde{\pi}_i^p - \tilde{\pi}^a = \sum_s w_{s,i}^p \pi_{s,i}^p - \sum_s w_s^a \pi_s^a$$
$$= \left\{\sum_s w_s^a \left(\pi_{s,i}^p - \pi_s^a\right)\right\} + \left\{\sum_s \left(w_{s,i}^p - w_s^a\right)\pi_s^a\right\} + \left\{\sum_s \left(w_{s,i}^p - w_s^a\right)\left(\pi_{s,i}^p - \pi_s^a\right)\right\}.$$

The first term captures the contribution of the manager's errors about category-specific inflation rates. The second term captures the contribution of the manager's errors about the weights applied to each category. The

final term captures the covariance between errors in weights and errors in ainflation rates across categories.

Results of this decomposition are reported in table 5. First, we report the average values of the error across all firms as well as the average values of each of the terms in the decomposition. The average manager error is 1.7 percentage points, implying that on average, managers overestimated the positive effect of these categories on aggregate inflation. Errors about category-level inflation rates contributed about 1.2 out of the 1.7 percentage points. Therefore, much of the average error can be explained by the fact that, on average, managers significantly overestimated the increase in food prices. The contribution of errors about weights is smaller, 0.8 percentage point out of the 1.7 percentage points, and comes primarily from the fact that managers overestimated the weights on housing and stock prices. The negative covariance term is driven largely by stock prices and food prices, in which respondents over(under)estimated the weights but under(over)estimated the inflation rate. These results suggest that managers' errors about recent inflation rates for specific categories of goods, particularly food prices, can account for much of their average misperceptions of aggregate inflation.

However, another feature worth noting in table 5 is that the cross-sectional standard deviations of inflation perceptions for most categories of goods are relatively low. In fact, for all but stock prices and gasoline prices (two very volatile price series), there is actually less disagreement among managers about inflation at the category level than there is about aggregate inflation. This suggests that disagreement about category-level inflation rates among managers is unlikely to account for the amount of disagreement that we observe in managers' beliefs about aggregate inflation rates. Consistent with this, we report in table 5 the cross-sectional standard deviations of each term in the decomposition of the errors above. Disagreement among managers about the weights assigned to different categories accounts for three times more of the dispersion in inflation errors than disagreement about category-specific inflation rates.

This last result suggests that the primary source of the large disagreement that we observe in managers' perceptions of recent inflation rates is differences in opinion about the relative importance of different categories of goods in the measurement of inflation.[13]

13. This result is not sensitive to the particular decomposition we used. For example, another decomposition is $\tilde{\pi}_i^p - \tilde{\pi}^a = \left\{\sum_s w_s^a \left(\pi_{s,i}^p - \pi_s^a\right)\right\} + \left\{\sum_s \left(w_{s,i}^p - w_s^a\right)\left(\pi_{s,i}^p - \tilde{\pi}_i^p\right)\right\} + \left\{\sum_s \tilde{\pi}_i^p \left(w_{s,i}^p - w_s^a\right)\right\}$, where the first term is the same as before, the second measures the

Table 6. Narrative Responses to How Managers Form Their Inflation Expectations[a]

"How do you typically form your inflation expectations?"	*Percent of managers* (1)	*Absolute backcast error*		*Inflation forecast*	
		Mean (2)	*SD* (3)	*Mean* (4)	*SD* (5)
Media	47	2.61	1.78	3.65	2.45
Meetings and discussions	10	3.10	1.52	4.00	1.22
Shopping experience	43	4.27	1.47	4.95	1.12
Prices of competitors and suppliers	10	4.50	2.17	5.20	2.39
Total	—	3.31	1.81	4.24	1.95

Source: Authors' survey of New Zealand firms.

a. Narrative responses may be classified into several categories. Of all the survey responses, only one response did not fit into any of the four groups indicated in the left column. For columns 1 and 4, the values are expressed as percents; all other values are expressed as percentage points..

IV.B. How Do Managers Form Inflation Expectations?

For managers to hold such different expectations about inflation, they must rely on different sources of information to form their expectations. We investigate the sources of managers' inflation expectations in two ways. First, in the fifth wave of the survey, we asked managers the following open-ended question:

> How do you typically form your inflation expectations?

Managers provided four general types of responses, as summarized in table 6.

First, 47 percent of respondents stated that they relied primarily on media sources for their information about inflation. These respondents tended to have much better information about recent inflation dynamics than others, with average absolute backcast errors being smaller by one percentage point on average; inflation forecast errors were lower as well. The second

extent to which agents place too much (too little) weight on categories of goods for which inflation is higher (lower) than average, and the third term captures potential errors from assigning too much or too little cumulative weights to all of the listed categories of goods (since the weights do not need to sum to one). This decomposition also implies that average errors primarily reflect errors in category-level inflation rates (since the first term in this decomposition is identical to ours). It similarly implies that the cross-sectional dispersion in inflation forecast errors is largely due to differences in beliefs about weights, since the cross-sectional standard deviation of the second term is almost three times that of the first term, about the same ratio as in our original decomposition. We are grateful to David Romer for suggesting this alternative way of decomposing forecast errors.

Table 7. Quantitative Responses to How Managers Form Their Inflation Expectations[a]

		Importance of information source			
	Share of managers (1)	*Family and friends (2)*	*Employees and colleagues (3)*	*Customers and suppliers (4)*	*Gas prices (5)*
All firms	1.00	2.12	1.88	1.68	4.14
If response is "very important" or "extremely important"					
Family and friends	0.06	4.67	4.00	1.67	4.00
Employees and colleagues	0.08	3.25	4.50	2.00	4.25
Customers and suppliers	0.04	1.00	1.50	4.50	4.00
Gas prices	0.76	2.24	1.82	1.74	4.53
Personal shopping experience	0.88	2.07	1.77	1.68	4.25
Media	0.54	2.00	1.81	1.59	4.07
Professional forecasts	0.20	2.00	2.10	1.30	3.90
Correlation matrix					
Family and friends		1.00			
Employees and colleagues		0.34	1.00		
Customers and suppliers		–0.13	0.15	1.00	
Gas prices		0.11	–0.14	–0.07	1.00
Personal shopping experience		–0.14	–0.27	–0.01	0.65
Government agencies		–0.10	0.09	0.31	0.15
Business associations, fairs		–0.11	0.12	0.31	–0.21
Media		–0.15	–0.06	–0.05	0.02
Professional forecasts		0.03	0.10	0.01	–0.14

Source: Authors' survey of New Zealand firms.

a. Respondents were asked to assess importance of various sources of information for their formation of inflation expectations. The scale runs from 1 ("not important") to 5 ("extremely important"). Columns 2–10 report average responses, collected in the fifth wave of the survey. Columns 11 and 12 report the average absolute backcast error (percentage point) and average inflation forecast (percent) of the respondents..

most common answer, accounting for 43 percent of respondents, was that they relied on their personal shopping experience to inform them about price changes. Many respondents emphasized housing prices as a particularly important source of information. In comparison with managers who relied on media, those who relied on their personal shopping experience tended to have larger errors about recent inflation dynamics and higher inflation forecasts. The remaining two answer categories account for much smaller shares of respondents, approximately 10 percent each. One answer is that managers discuss inflation with coworkers or family members ("meetings and discussions"). Respondents who chose this answer tended to have relatively good information about inflation. The other answer states that they rely on their competitors' or suppliers' prices to make inferences about aggregate inflation. This group had the largest average errors about

Personal shopping experience (6)	*Government agencies (7)*	*Business associations, fairs (8)*	*Media (9)*	*Professional forecasts (10)*	*Average absolute backcast error (11)*	*Average inflation forecast (12)*
4.42	1.74	1.82	3.54	2.42	3.27	4.18
3.67	1.67	2.00	2.33	2.67	3.37	4.67
4.00	2.00	1.75	2.75	2.75	2.95	3.75
4.50	2.50	3.00	4.00	2.50	2.70	4.50
4.63	1.79	1.76	3.50	2.42	3.03	3.95
4.64	1.77	1.86	3.64	2.48	3.14	4.11
4.44	1.70	1.81	4.81	2.96	2.42	3.59
4.30	1.30	1.50	4.80	4.40	1.65	2.20
1.00						
0.09	1.00					
–0.07	0.38	1.00				
0.21	–0.05	–0.02	1.00			
–0.11	–0.29	–0.07	0.48	1.00		

recent inflation. While there are differences in beliefs within each group, these results do suggest that the average effect of the main source of information for inflation can be very large: the average difference in inflation backcast errors can be as large as 1.6 percentage points.

To investigate this insight in more detail, we also asked respondents to rank the importance of nine different sources of information to them in forming their inflation expectations on a scale of 1 (lowest) to 5 (highest). The specific sources were (i) family and friends, (ii) employees and colleagues, (iii) customers and suppliers, (iv) gas prices, (v) personal shopping experience, (vi) government agencies, (vii) business associations, chambers of commerce, and trade fairs, (viii) media (television, newspapers, and so on), and (ix) professional forecasts. The average ranks given by managers are listed in the first row of table 7.

Consistent with the narrative responses, two of the most highly ranked categories were personal shopping experience and media. In addition, these results highlight the particular importance of gasoline prices as a reference point to managers for making inferences about broader inflation movements: 76 percent of managers rank gasoline prices as very important or extremely important (rank of 4 or 5) to them in forming their inflation expectations. This is consistent with the argument of Coibion and Gorodnichenko (2015b) that households place a disproportionate amount of weight on oil or gasoline prices in forming their inflation expectations.

Table 7 also presents the average ranks assigned to categories by managers who rated specific categories as very or extremely important to them, the average backcast errors and inflation forecasts for these groups of firms, and the correlation matrix of ranks given by respondents to different sources of information. These jointly yield several results. First, managers who rated professional forecasters as very or extremely important to them had much more accurate perceptions and forecasts of inflation, on average, than other managers. They also tended to rely on media and on employees or colleagues more than others did. However, only 20 percent of managers rate professional forecasts as being this important. Second, as we found with the narrative questions, managers who rated media as very or extremely important also had more accurate information about inflation than others, on average. Approximately 54 percent of managers rated media as very or extremely important to them.

Third, and perhaps most strikingly, 88 percent of managers rated their personal shopping experience as very or extremely important. Even among those respondents who utilized professional forecasts and media reports extensively, the average ratings on personal shopping experience continue to be very high. This suggests that personal shopping experience is a more important source of information to most managers than might have been implied by the narrative approach, in which managers seem to generally report only their first source of information. Because shopping experiences are likely to be so heterogeneous across agents, both in the prices managers pay and the share of expenditures they allocate to different categories of goods, the systematic importance of an individual's shopping experience suggests a very natural source for the wide variation that we observe in beliefs about inflation across managers.

IV.C. Do Managers' Inflation Expectations Affect Their Decisions?

Does it make any difference to managers' decisions whether they expect inflation to be 1 percent or 5 percent? One way to answer this would

Table 8. Narrative Responses to How Managers Use Their Inflation Expectations[a]

"How do you typically use your inflation expectations?"	*Percent of managers* (1)	*Absolute backcast error*		*Inflation forecast*	
		Mean (2)	*SD* (3)	*Mean* (4)	*SD* (5)
Personal use	67	3.35	1.78	4.15	1.82
Price-setting decisions	12	2.20	1.64	3.33	2.88
Wage-setting decisions	14	4.27	2.07	4.43	1.51
Investment decisions	18	3.03	1.58	5.11	2.15
Total	—	3.26	1.84	4.18	2.01

Source: Authors' survey of New Zealand firms.

a. Narrative responses may be classified into several categories. For all columns, the values are expressed as percentage points.

be to compare the decisions of different managers who hold different inflation expectations, but that approach would require us to be able to control for a wide array of factors relevant to each manager's decision. Instead, we used hypothetical questions to investigate how managers would respond to changes in their expectations. In the fifth wave of the survey, we asked managers an open-ended question about how inflation expectations affect their decisions. They were asked:

> How do you typically use your inflation expectations?

Managers gave four general types of answers, as summarized in table 8. Twelve percent of managers responded that inflation expectations mattered for their pricing decisions, 14 percent mentioned their wage-setting decisions, and 18 percent specified their investment decisions. Strikingly, two-thirds of managers reported that the primary use of their inflation expectations was for their own personal use, in terms of consumption and savings. Hence, similarly to how most managers relied on their personal experience as consumers to inform them about aggregate inflation developments, we now find that managers report that the primary use of their inflation expectations is for their personal decisionmaking rather than for the economic decisions of the firm. This provides a novel justification for why managers' inflation expectations resemble those of households in so many ways: their primary use is for their personal consumption and saving decisions.

Of course, the fact that managers respond to an open-ended question by saying that they use their inflation expectations mainly for their own

consumption and saving decisions, as did the majority of the respondents, does not imply that these managers do not use their inflation expectations at all for their business decisions, only that the latter are secondary to their use for personal decisions. In the third wave of the survey, we asked managers more restrictive hypothetical questions, focusing separately on prices, wages, employment, and investment decisions:

> If you thought overall prices in the economy over the next 12 months were going to rise by more than what you are currently forecasting, would you be more likely to [increase/decrease/no change] your [prices/employment/investment/wages]?

Only 25 percent of managers reported that higher inflation expectations would have no effect on any of their economic decisions, whereas 75 percent responded along at least one margin. Hence, most managers do report that changes in their inflation expectations would induce them to alter some of the economic choices made by their firm.

The responses for each individual variable are presented in table 9. Approximately 35 percent of firm managers report that they would charge higher prices, whereas between 25 percent and 30 percent of firms report that they would raise wages, employment, or investment. Almost no managers would decrease any of these variables. Table 9 also considers pairs of answers across variables. Few managers would adjust along more than one margin in these pairings: Just 13 percent of managers would raise prices and wages, 8 percent of managers would raise employment and wages, and 9 percent of managers would raise investment and employment. Instead, the majority of firms would pursue adjustment along a single margin, although the specific margin they choose is difficult to predict. The key result from this survey question is that most firm managers appear to treat their inflation expectations as one of the inputs into their decision process. This implies that *if* policymakers can change managers' inflation expectations, then one should expect some economic repercussions through pricing, wage, employment, and investment decisions to occur.

V. Results for the United States

While New Zealand presents a particularly interesting case study for inflation expectations given its long experience with inflation targeting, one would like to know whether the characteristics documented in the previous section extend to other countries. In this section, we revisit whether the inflation expectations of the U.S. public also appear

Table 9. How Managers Would Respond to Higher Inflation Expectations

	All responses[a]		
Variable[a]	*Increase*	*No change*	*Decrease*
Price	0.354	0.604	0.042
Wage	0.253	0.727	0.019
Employment	0.274	0.715	0.011
Investment	0.293	0.694	0.013
Wage vs. price[b]	*Wage response*		
Price response	*Increase*	*No change*	*Decrease*
Increase	0.130	0.217	0.007
No change	0.114	0.480	0.011
Decrease	0.009	0.030	0.002
Employment vs. wage[b]	*Employment response*		
Wage response	*Increase*	*No change*	*Decrease*
Increase	0.082	0.167	0.004
No change	0.184	0.536	0.007
Decrease	0.007	0.012	0.000
Employment vs. investment[b]	*Employment response*		
Investment response	*Increase*	*No change*	*Decrease*
Increase	0.089	0.200	0.004
No change	0.182	0.504	0.007
Decrease	0.002	0.011	0.000

Source: Authors' survey of New Zealand firms.

a. Reports shares of managers responding "increase," "decrease," or "no change" to the survey question "If you thought overall prices in the economy over the next 12 months were going to rise by more than what you are currently forecasting, would you be more likely to [increase/decrease/no change] your [prices/employment/investment/wages]?"

b. Reports breakdown of how managers responded to the named pair of actions in the questions.

unanchored, and whether the U.S. public displays the same knowledge of monetary policy as firm managers in New Zealand. Because no broad quantitative survey of firm managers exists for the United States, we focus on the University of Michigan's Survey of Consumers, the Federal Reserve Bank of New York's Survey of Consumer Expectations, and the Federal Reserve Bank of Philadelphia's Survey of Professional

Forecasters; following Binder (2015), we use additional polling data as well when available.[14]

V.A. Are the U.S. Public's Inflation Expectations Anchored?

First, we revisit the five predicted characteristics of anchored expectations presented in section II that failed to be present in New Zealand. First, we examine whether average beliefs are close to the inflation target of the central bank. Table 10 shows that both short-term and long-term inflation expectations of households hover between 3.1 and 4.4 percent, well above the Federal Reserve's 2 percent inflation target. In contrast, professional forecasters predict inflation to be close to the official target at all horizons. The actual rate of inflation over the 2013Q4–2014Q4 period was less than 2.4 percent. Strikingly, the magnitudes are similar to those predicted by New Zealand's counterparts.[15]

Second, we explore whether economic agents in the United States disagree about the future course of inflation. We find that while the cross-sectional dispersion of projections made by professional forecasters is small (approximately 0.4 percentage point), the dispersion is an order of magnitude larger for households (approximately 4 percentage points at short horizons and somewhat smaller for longer horizons). Hence, along this metric as well, the U.S. public's inflation expectations look no more anchored than those of the New Zealand public.

Third, we investigate how much confidence agents have in their forecasts. The Survey of Consumer Expectations asks respondents to assign probabilities to 10 inflation bins. Using this information, we can calculate the implied standard deviation, a measure of forecast uncertainty. We find that although there is considerable heterogeneity across U.S. consumers, they are generally very uncertain in their predictions: The mean standard deviation is approximately 3 percentage points, which is comparable to

14. The polling data reported here were obtained from searches of the iPOLL Databank and other resources provided by the Roper Center for Public Opinion Research, University of Connecticut.

15. Neither the Michigan Survey of Consumers nor the Survey of Consumer Expectations asks people to report current or past inflation. A poll by the Pew Research Center for the People and the Press asked the public about whether recent inflation data were closer to 1, 5, 10, or 20 percent. The most common answer, with 49 percent of responses, was "don't know/refused to answer" and the mean response was 7.4 percent, well above the actual inflation rate of approximately 1 percent. (See Binder [2015] for more details.) Thus, similar to consumers in New Zealand, consumers in the United States appear to have a perception of inflation well above actual figures.

Table 10. Inflation Forecasts of Firm Managers and Others, United States, 2013Q4–2014Q4[a]

					Households			
			Survey of Professional Forecasters		Michigan Survey of Consumers		Survey of Consumer Expectations	
Survey date	Recent data (1)	Central bank (2)	Mean (3)	SD (4)	Mean (5)	SD (6)	Mean[b] (7)	SD (8)
1-year-ahead inflation forecast								
2013Q4	1.4	1.5	1.9	0.5	3.7	3.8	4.4	4.5
2014Q1	2.1	1.8	1.9	0.5	4.1	4.0	4.2	4.3
2014Q2	2.4	1.8	1.9	0.5	4.0	3.5	4.3	4.4
2014Q3	1.2	1.8	2.1	0.5	3.9	3.7	4.2	4.3
2014Q4	–0.9	1.3	1.9	0.5	3.1	3.6	n.a.	n.a.
Long-term inflation forecast[c]								
2013Q4		2.0	2.1	0.4	3.4	3.2	4.4	4.5
2014Q1		2.0	2.1	0.4	3.4	3.1	4.1	4.4
2014Q2		2.0	2.2	0.3	3.5	3.0	4.3	4.6
2014Q3		2.0	2.2	0.4	3.3	2.9	4.2	4.5
2014Q4		2.0	2.1	0.3	3.1	2.7	n.a.	n.a.

Sources: Economic Projections of Federal Reserve Board Members and Federal Reserve Bank Presidents, Survey of Professional Forecasters, Michigan Survey of Consumers, and Survey of Consumer Expectations.

a. Reports actual inflation, inflation forecasts of the Federal Reserve System, mean forecasts from surveys, and the cross-sectional standard deviations of these forecasts (columns 3–8). Standard deviations are expressed as percentage points, and all other values are expressed as percents.

b. Mean inflation forecast implied by distribution of respondents' expectations about future inflation.

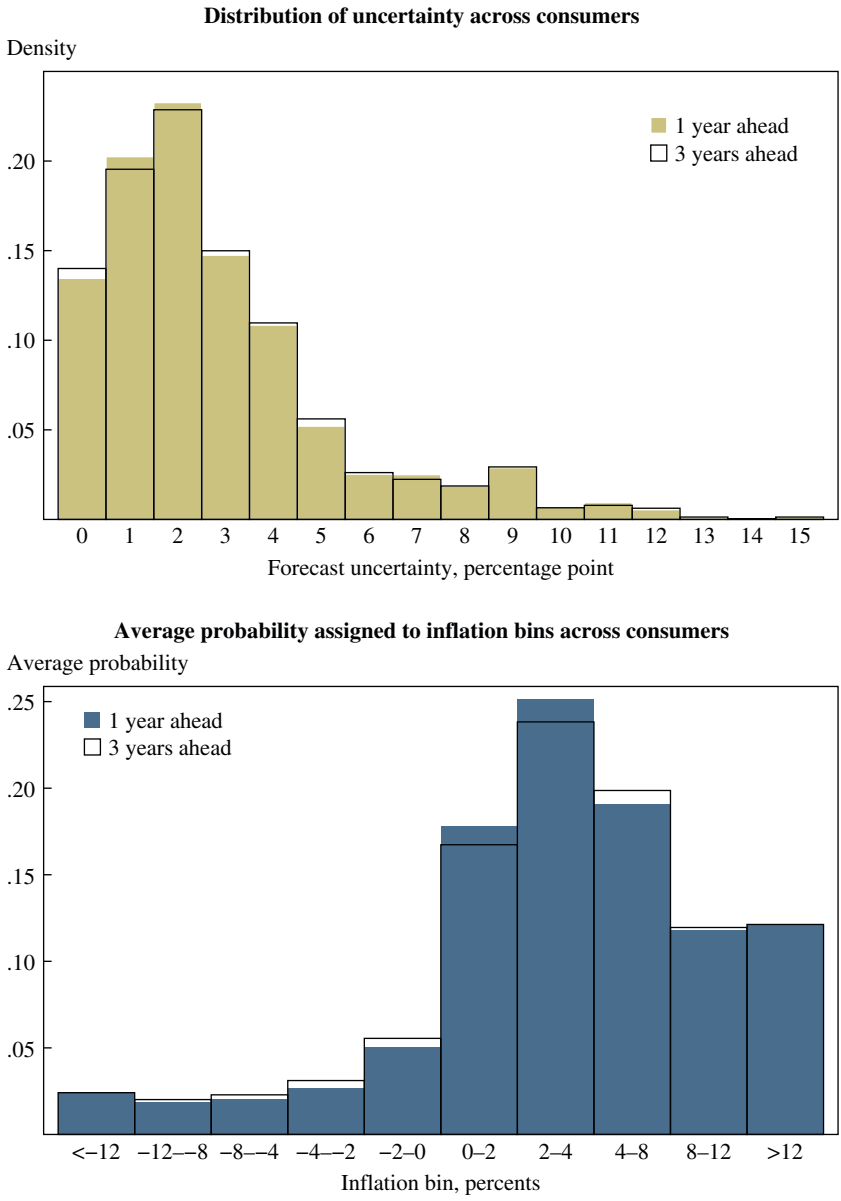
c. The horizons of the long-term forecasts are 5 years for the Survey of Professional Forecasters and the Michigan Survey of Consumers and 3 years for the Survey of Consumer Expectations. For the projections of the central bank, we use "longer run" projections in the lower panel and the middle of the central tendency range for CPI for the next calendar year.

the mean forecast of consumers; see the upper panel of figure 7. The lower panel of figure 7 also shows that, on average, consumers assign more than 50 percent probability to inflation in the United States being greater than 4 percent over the next 12 months and over the next three years.

Other survey evidence corroborates this result. For example, Binder (2015) compiles extensive poll evidence that consumers have little confidence in low and stable inflation. According to the Retirement Confidence Survey,[16] in years 2012 and 2013, only 6 percent of respondents claimed to be very confident that inflation will remain moderate over the next

16. Conducted by the Employee Benefit Research Institute.

Figure 7. Uncertainty in Consumers' Inflation Forecasts in the U.S.[a]



Source: Authors' survey of New Zealand firms.

a. The survey question asks managers to assign probabilities to nine bins. For each respondent, we construct a measure of uncertainty as the standard deviation of the reported distribution.

10 years and will average no more than 4 percent in the next 10 years. At the same time, approximately 60 percent of respondents stated that they are either not at all confident or not too confident in these outcomes. In addition, Binder (2014) shows that confidence in forecasts as measured by the rounding of point predictions has shown little change since the mid-1980s. These patterns are similar to those observed in New Zealand.

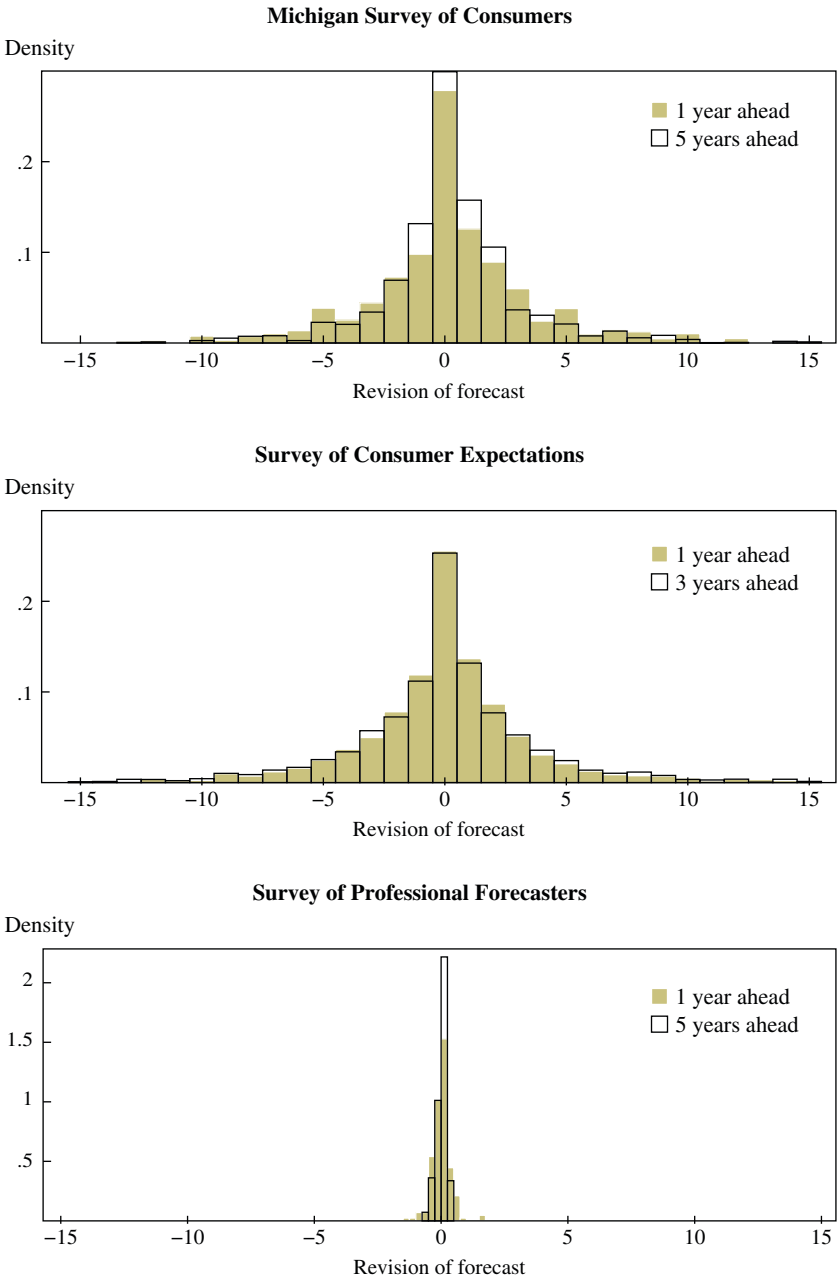
Fourth, we consider the size of revisions in inflation forecasts. Figure 8 presents both 1-year-ahead and 3- or 5-years-ahead inflation forecasts. As was the case with consumers and firm managers in New Zealand, the average absolute size of revisions is very large for U.S. consumers. Households frequently revise their inflation forecasts by as much as five percentage points or more. Binder (2015) shows that the share of Michigan Survey of Consumers respondents who revise their inflation forecasts has been between 70 and 80 percent since the early 1980s. In contrast, revisions in inflation forecasts for professional forecasters are much smaller and resemble what one would expect to see under anchored expectations, in sharp contrast to those of U.S. households.

Finally, we regress long-run inflation forecasts on short-term inflation forecasts. We do so both in levels (as with the New Zealand data) and using revisions in both short-run and long-run forecasts at the individual level, since this type of panel data is available for the United States in the Michigan Survey of Consumers, the Survey of Consumer Expectations, and the Survey of Professional Forecasters. As in New Zealand, in the United States long-term forecasts are highly sensitive to movements in short-term forecasts (table 11). This sensitivity remains large even after controlling for consumer fixed effects. Furthermore, controlling for outliers tends to yield even higher estimates of the sensitivity: A 1-percentage-point increase in the short-term inflation forecast can be associated with as much as a 1-percentage-point increase in the long-term inflation forecast. These sensitivities are broadly in line with the sensitivity estimated for managers in New Zealand. On the other hand, the sensitivity is much weaker for professional forecasters.

V.B. What Does the U.S. Public Know about Monetary Policy?

Given the apparent lack of anchoring in the inflation expectations of the U.S. public, we would like to know to what extent this reflects shortages in knowledge about monetary policy and recent inflation dynamics as opposed to a lack of credibility on the part of the Federal Reserve. As with New Zealand, we consider how well the public knows who chairs the Federal Reserve Board as a simple measure of how informed the public is

Figure 8. Forecast Revisions by U.S. Households and Professional Forecasters[a]



a. In each panel, shaded bars are for the distribution of revisions in long-term (3- or 5-years-ahead) inflation forecasts. Nonshaded bars are for the distribution of revisions in short-term (1-year-ahead) inflation forecasts.

Table 11. Correlation between Short-Term and Long-Term Inflation Forecasts, United States[a]

	Level			*3-month revision*			*6-month revision*		
	OLS[b] *(1)*	*WGT*[c] *(2)*	*RREG*[d] *(3)*	*OLS*[b] *(4)*	*WGT*[c] *(5)*	*RREG*[d] *(6)*	*OLS*[b] *(7)*	*WGT*[c] *(8)*	*RREG*[d] *(9)*
Survey of Consumer Expectations, truncated sample, 3-years-ahead forecast[e,f]									
1-year-ahead forecast	0.672***	0.671***	0.932***	0.477***	0.505***	0.681***	0.507***	0.557***	0.723***
	(0.012)	(0.017)	(0.003)	(0.026)	(0.030)	(0.005)	(0.037)	(0.043)	(0.008)
No. of observations	17,854	17,849	17,854	8,288	8,288	8,288	3,513	3,513	3,513
R^2	0.392	0.371	0.879	0.203	0.221	0.664	0.211	0.258	0.701
Survey of Consumer Expectations, full sample, 3-years-ahead forecast[e]									
1-year-ahead forecast	0.562***	0.586***	0.984***	0.399***	0.412***	0.706***	0.386***	0.424***	0.857***
	(0.022)	(0.035)	(0.001)	(0.034)	(0.049)	(0.002)	(0.046)	(0.070)	(0.003)
No. of observations	20,573	20,565	20,573	10,031	10,030	10,031	4,275	4,275	4,275
R^2	0.295	0.299	0.984	0.141	0.143	0.944	0.135	0.136	0.960
Survey of Consumer Expectations, implied point prediction, 3-years-ahead forecast[g]									
1-year-ahead forecast	0.787***	0.783***	0.975***	0.643***	0.623***	0.815***	0.647***	0.662***	0.829***
	(0.009)	(0.014)	(0.002)	(0.019)	(0.027)	(0.005)	(0.024)	(0.034)	(0.007)
No. of observations	20,567	20,559	20,567	10,047	10,046	10,047	4,285	4,285	4,285
R^2	0.571	0.571	0.936	0.350	0.339	0.756	0.373	0.402	0.781
Michigan Survey of Consumers, truncated, 5-years-ahead forecast[f]									
1-year-ahead forecast	0.442***	0.444***	0.420***				0.349***	0.355***	0.285***
	(0.012)	(0.014)	(0.005)				(0.027)	(0.031)	(0.011)
No. of observations	10,443	10,320	10,443				2,371	2,352	2,371
R^2	0.280	0.281	0.428				0.159	0.166	0.231

(*continued on next page*)

Table 11. Correlation between Short-Term and Long-Term Inflation Forecasts, United States[a] (*Continued*)

	Level			*3-month revision*			*6-month revision*		
	OLS[b] (1)	*WGT*[c] (2)	*RREG*[d] (3)	*OLS*[b] (4)	*WGT*[c] (5)	*RREG*[d] (6)	*OLS*[b] (7)	*WGT*[c] (8)	*RREG*[d] (9)
Michigan Survey of Consumers, full sample, 5-years-ahead forecast									
1-year-ahead forecast	0.442***	0.445***	0.411***				0.336***	0.350***	0.283***
	(0.012)	(0.013)	(0.004)				(0.026)	(0.029)	(0.009)
No. of observations	10,588	10,464	10,588				2,440	2,420	2,440
R^2	0.301	0.306	0.470				0.171	0.185	0.281
Survey of Professional Forecasters, 5-years-ahead forecast									
1-year-ahead forecast	0.565***		0.598***	0.169***		0.146***	0.191***		0.163***
	(0.035)		(0.028)	(0.034)		(0.029)	(0.035)		(0.032)
No. of observations	286		286	244		244	238		238
R^2	0.572		0.619	0.105		0.095	0.125		0.097
Survey of Professional Forecasters, 10-years-ahead forecast									
1-year-ahead forecast	0.474**		0.468**	0.057**		0.050**	0.056*		0.065**
	(0.037)		(0.033)	(0.028)		(0.021)	(0.029)		(0.026)
No. of observations	285		285	244		244	237		237
R^2	0.398		0.408	0.014		0.024	0.012		0.025

a. Reports results of regressing long-term inflation forecasts (or forecast revisions) on short-term inflation forecasts (or forecast revisions). Sample period covers 2013Q–2014Q4. Standard errors in parentheses for columns 3, 6, and 9; robust standard errors in parentheses for the remaining columns. Statistical significance at the ***1 percent, **5 percent, and *10 percent levels.

b. Does not use sampling weights in estimation.

c. Uses sampling weights in estimation.

d. Uses Huber robust regression to downweight the importance of outliers and influential observations.

e. Uses point predictions reported in the Survey of Consumer Expectations.

f. Sample includes only respondents reporting inflation forecasts less than 15 percent in absolute value.

g. Uses mean predicted inflation implied by the reported probability distribution in the Survey of Consumer Expectations.

about recent monetary policy actions and discussions. Binder (2015) documents that the public is largely unaware of who the chair of the Federal Reserve is. In a number of polls asking the public to pick the name of the chair from four options, between 20 and 50 percent of respondents simply refuse to answer these multiple choice questions, while approximately one-third tend to pick the correct answer out of four names. The lowest share of correct answers in the September 2014 poll was for Janet Yellen (24 percent), but this was only 6 months after she became chair, which may explain her lower name recognition. These relatively low shares of correct answers are very close to those observed in the survey in New Zealand, suggesting a similar lack of awareness of the leadership of the central bank in both countries among much of the broader public.

Just as we found that the general managers of firms in New Zealand did not seek out information about monetary policy, we can explore whether the U.S. public tries to access information about monetary policy directly from the source through social media. Table 12 presents numbers of Facebook and Twitter followers for each of the Federal Reserve Banks as well as for the Federal Reserve System as a whole. As of June 30, 2015, the entire Federal Reserve System had 702,955 followers on Twitter. For comparison, the U.S. State Department had nearly twice as many followers, the FBI had roughly 470,000 more, and the CIA had roughly 120,000 more. In fact, the voice of the entire Federal Reserve System on Twitter is singlehandedly dwarfed by that of Paul Krugman, who has almost twice as many followers. Even former congressman Ron Paul and Senator Rand Paul each has almost as many followers as the entire Federal Reserve System. Binder (2015) documents similar evidence.

Data from Google Trends also allow us to verify the extent to which the U.S. public seeks out information about macroeconomic conditions online. Figure 9 plots the volume of online Google searches in the United States since 2004 for the macroeconomic variables "GDP," "inflation," and "unemployment rate" (relative to total searches). For comparison, we also plot the volume of searches for "puppies" (relative to total searches), which is roughly three to four times as high. Strikingly, with the onset of the Great Recession, there is only a slight increase in the volume of searches for macroeconomic variables, and it reverses in 2009. Binder (2015) finds a very similar pattern for an alternative but related set of searches.

But just as in New Zealand, the fact that much of the U.S. public does not actively seek out information about macroeconomic conditions or monetary policy does not mean that it necessarily views this information as unimportant. A poll from CNN and the Opinion Research Corporation in

Table 12. Social Media Following of the Federal Reserve and Others, June 2015[a]

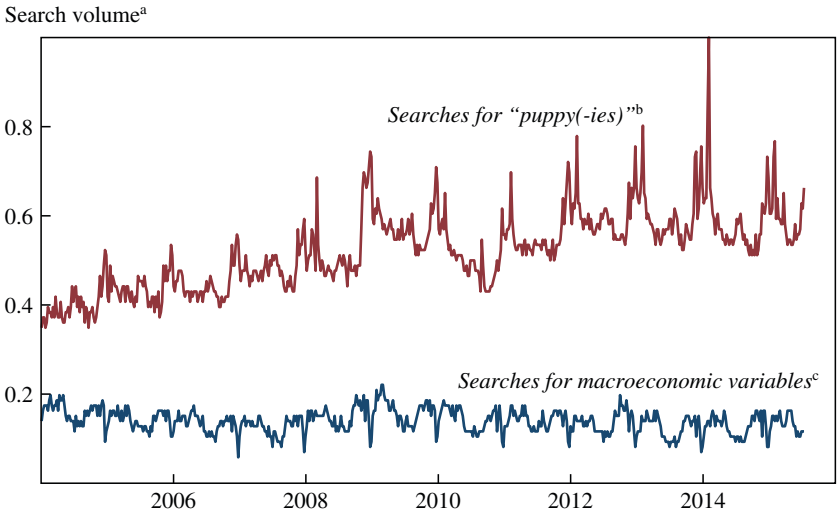
Institution or person with following	*Facebook "likes"*	*Twitter followers*
Federal Reserve	—	268,258
Federal Reserve Bank of Atlanta	1,854	33,932
Federal Reserve Bank of Boston	—	31,947
Federal Reserve Bank of Chicago	—	45,640
Federal Reserve Bank of Cleveland	1,506	26,920
Federal Reserve Bank of Dallas	—	31,030
Federal Reserve Bank of Kansas City	—	14,385
Federal Reserve Bank of Minneapolis	811	29,261
Federal Reserve Bank of New York	7,787	86,991
Federal Reserve Bank of Philadelphia	1,254	34,767
Federal Reserve Bank of Richmond	432	20,895
Federal Reserve Bank of St. Louis	4,416	49,411
Federal Reserve Bank of San Francisco	3,736	29,518
Total	21,796	702,955
U.S. Department of Energy	69,084	249,663
U.S. Department of Homeland Security	320,261	663,009
U.S. Department of Justice	183,507	880,396
U.S. Department of State	1,103,944	1,408,592
U.S. Department of the Treasury	41,613	275,636
Total	1,718,409	3,477,296
CIA	440,716	818,493
FBI	1,235,614	1,170,177
Paul Krugman	—	1,374,547
Ron Paul	1,287,106	549,344
Rand Paul	2,024,694	637,037

a. Facebook and Twitter pages were accessed on June 30, 2015.

2011 asked respondents to evaluate how important different issues would be to their vote in the 2012 presidential elections, and found that 68 percent of respondents rated inflation as extremely important or very important to them. This combined share of importance put inflation just below terrorism and taxes in importance to the public and above the war in Afghanistan, illegal immigration, guns, the situation in Libya at the time, abortion, and gay marriage (see figure 10), despite the fact that the president's influence on inflation is much more limited than on any of these other issues. In a similar spirit, Binder (2015) shows that there has been an increasing trend in the share of people claiming that they are hurt by inflation.

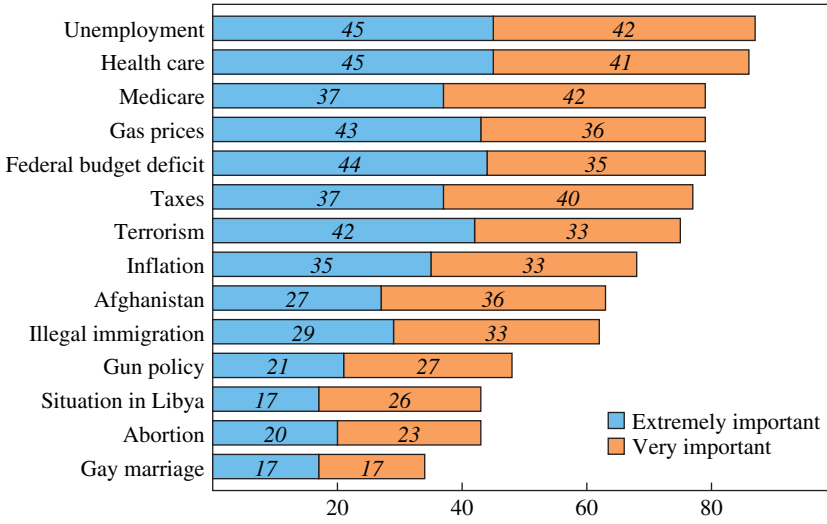
A likely reason for the apparent lack of active interest in monetary policy may be the view that it is a difficult topic to understand. The history of monetary policy communications is, of course, not one of openness

Figure 9. U.S. Google Searches for Macroeconomic Variables vs. "Puppies," January 2004–July 2015



Source: Google Trends, accessed July 12, 2015.
a. Volume of online Google searches, relative to the highest point on the graph (normalized to 1).
b. The two search terms are "puppy" and "puppies." The search volumes are aggregated together.
c. The three macroeconomic variable search terms are "GDP," "inflation," and "unemployment rate." The search volumes are aggregated together.

Figure 10. Relative Importance of Inflation as an Issue to the U.S. Public, 2011[a]



Source: June 2011 CNN and Opinion Research Corporation Poll.
a. The poll question asked, "How important will each of the following issues be to your vote for President next year (2012)?" Values are expressed as percents.

with the public. While central banks like the Federal Reserve have over the last two decades become increasingly communicative with the public, this increase in communication might not necessarily have helped matters much for the general public. Rubén Hernández-Murillo and Hannah Shell (2014), for example, analyze the complexity of Federal Open Market Committee (FOMC) statements since the early 1990s and find that the length and reading level associated with these statements has increased significantly over time. For example, in the mid-1990s, a typical FOMC statement was approximately 100 words long and required a 12th-grade reading level to understand. By 2014, the length of the statement was six times as long and the reading level was effectively that of a Ph.D.

Not surprisingly, when members of the U.S. public are asked how well they understand monetary policy, they tend to express hesitation. In a July 2014 poll, the Associated Press asked respondents to indicate how easy or hard they felt it was to understand Federal Reserve policy on interest rates. While 27 percent of respondents claimed it was very easy or somewhat easy, 70 percent expressed difficulty understanding monetary policy, and a quarter of respondents said it was "very hard." Binder (2015) presents additional evidence documenting the complexity of the Federal Reserve's communication. In short, despite dramatic changes in the communications strategy of the Federal Reserve over the last 20 years, the U.S. general public appears to remain profoundly uncertain about what exactly the Federal Reserve does.

VI. Conclusion

After 25 years of largely successful inflation targeting in New Zealand, the inflation expectations of households and managers there do not appear particularly well anchored. Managers of firms disagree dramatically about recent and future inflation levels, even at long horizons, and many are poorly informed about the RBNZ's inflation target. Most managers appear to rely to a large extent on their personal shopping experience to make inferences about aggregate inflation, and they are particularly sensitive to gasoline prices, as is the case with households in the United States. Indeed, along most metrics, the expectations of managers are much more similar to the expectations of households than to those of professional forecasters. Since neither group appears to be well informed about monetary policy overall, this suggests that changes over the last 20 years in central banks' communications strategies have not had the desired effects on this segment of the population, even if they may have been more effective with

professionals or financial market participants. In short, while this state of affairs may be better than what New Zealand had before adopting inflation targeting (we do not have data on inflation expectations covering that period), in absolute terms the properties of inflation expectations and perceptions of inflation appear to score poorly along the basic metrics of how anchored expectations are.

The lack of awareness by the general public, including firm managers, of the actions and objectives of monetary policymakers in a low-inflation environment is particularly problematic in periods when central bankers are seeking to affect inflation expectations through forward guidance. Because such policies are designed to have real effects precisely by generating changes in agents' inflation expectations and therefore in their perceived real interest rates, the fact that the public may largely be unaware of the policies or of their implications for aggregate prices implies that their effects will most likely be limited, or at least much less than predicted by models with full-information rational-expectations agents. At least some central bankers are aware of the limited power of their promises.[17]

Coming to terms with these limited, heterogeneous information sets on the part of households and firms, not just regarding knowledge of contemporaneous economic conditions but also regarding knowledge of the central bank's objectives, will be challenging from a modeling point of view. For example, Coibion and Gorodnichenko (2011) show that the strategic interaction of firms with different practices for pricing and acquisition of information has important implications for macroeconomic dynamics and policy design. However, this and related research abstract from heterogeneity and the interaction of expectations across firms, households, professional forecasters, financial markets, and the central bank. Indeed, most macroeconomic models do not include financial markets or professional forecasters, so exactly how these arguably better-informed agents influence macroeconomic outcomes is not well understood.

Future work should shed new light on the optimal way for central banks to release information so it is useful for not just professional forecasters and financial market analysts but also the broader public. Only then may inflation targeting finally achieve its full promise.

17. Laurence Meyer (2004, p. 211) notes, "Greenspan believes that inflation expectations can best be anchored by a history of having achieved price stability rather than by a mere promise to do so."

ACKNOWLEDGMENTS We are grateful to Rochelle Barrow, Graham Howard, and Katrina Young for sharing data from the survey of households in New Zealand. We thank Alan Blinder, Lars Svensson, and the editors for comments on an earlier version of the paper. Gorodnichenko thanks the National Science Foundation for financial support.

References

Alvarez, Fernando E., Francesco Lippi, and Luigi Paciello. 2011. "Optimal Price Setting with Observation and Menu Costs." *Quarterly Journal of Economics* 126, no. 4: 1909–60.

Andrade, Philippe, and Hervé Le Bihan. 2013. "Inattentive Professional Forecasters." *Journal of Monetary Economics* 60, no. 8: 967–82.

Ball, Laurence, and Sandeep Mazumder. 2014. "A Phillips Curve with Anchored Expectations and Short-Term Unemployment." Working Paper no. 20715. Cambridge, Mass.: National Bureau of Economic Research.

Bernanke, Ben S. 2013. "Communication and Monetary Policy." Speech at the National Economists Club Annual Dinner, Herbert Stein Memorial Lecture, November 19. Washington. http://www.federalreserve.gov/newsevents/speech/bernanke20131119a.htm

Binder, Carola. 2014. "Measuring Uncertainty Based on Rounding: New Method and Application to Inflation Expectations." Working Paper. https://sites.google.com/site/carolabindereconomics/research

———. 2015. "Fed Speak on Main Street." Working Paper. https://sites.google.com/site/carolabindereconomics/research

Blinder, Alan S., Elie R. D. Canetti, David E. Lebow, and Jeremy B. Rudd. 1998. *Asking About Prices: A New Approach to Understanding Price Stickiness.* New York: Russell Sage Foundation.

Blinder, Alan S., Michael Ehrmann, Marcel Fratzscher, Jakob De Haan, and David-Jan Jansen. 2008. "Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence." *Journal of Economic Literature* 46, no. 4: 910–45.

Carroll, Christopher D. 2003. "Macroeconomic Expectations of Households and Professional Forecasters." *Quarterly Journal of Economics* 118, no. 1: 269–98.

Coibion, Olivier, and Yuriy Gorodnichenko. 2011. "Strategic Interaction among Heterogeneous Price-Setters in an Estimated DSGE Model." *Review of Economics and Statistics* 93, no. 3: 920–40.

———. 2012. "What Can Survey Forecasts Tell Us about Informational Rigidities?" *Journal of Political Economy* 120, no. 1: 116–59.

———. 2015a. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review* 105, no. 8: 2644–78.

———. 2015b. "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 Saten Kumar. 2015. "How Do Firms Form Their Expectations? New Survey Evidence." Working Paper no. 21092. Cambridge, Mass.: National Bureau of Economic Research.

Gorodnichenko, Yuriy. 2008. "Endogenous Information, Menu Costs and Inflation Persistence." Working Paper no. 14184. Cambridge, Mass.: National Bureau of Economic Research.

Hernández-Murillo, Rubén, and Hannah G. Shell. 2014. "The Rising Complexity of the FOMC Statement." *Economic Synopses,* no. 23.

Lucas, Robert E., Jr. 1972. "Expectations and the Neutrality of Money." *Journal of Economic Theory* 4, no. 2: 103–24.

Mackowiak, Bartosz, and Mirko Wiederholt. 2009. "Optimal Sticky Prices under Rational Inattention." *American Economic Review* 99, no. 3: 769–803.

Mankiw, N. Gregory, and Ricardo Reis. 2002. "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *Quarterly Journal of Economics* 117, no. 4: 1295–328.

Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers. 2004. "Disagreement about Inflation Expectations." *NBER Macroeconomics Annual* 18: 209–48.

Meyer, Laurence H. 2004. *A Term at the Fed: An Insider's View.* New York: HarperCollins.

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50, no. 3: 665–90.

Solow, Robert M. 1979. "What We Know and Don't Know about Inflation." *Technology Review* 81, no. 3: 30–47.

Svensson, Lars E. O. 2015. "Forward Guidance." *International Journal of Central Banking* 11, suppl. 1: 19–64.

Woodford, Michael. 2003. "Imperfect Common Knowledge and the Effects of Monetary Policy." In *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund Phelps,* edited by Philippe Aghion, Roman Frydman, Joseph Stiglitz, and Michael Woodford. Princeton University Press.

Comments and Discussion

COMMENT BY

ALAN S. BLINDER To begin with, I like this paper a lot—not because of a burning desire to learn about (mostly small) business managers in New Zealand, but because there are many reasons to believe, both in this paper and in Carola Binder's work, that the findings can be generalized. Saten Kumar, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko's work here has the strong ring of truth.

I find the paper valuable for three main reasons. First, it reminds us that most people are not obsessed about the central bank; as the authors note, they would rather watch puppies on YouTube. Second, it shows that expectations of inflation vary greatly across individuals, which raises the question of what the symbol $E(\pi)$ means in macro models. Third, as their title indicates, it shows that inflationary expectations are not well anchored in New Zealand, despite apparently good reasons why they should be. I will take up these three reasons in turn.

One way to state the paper's central finding is that small business managers in New Zealand think like ordinary people, not like the tiny minority of the human race (including virtually all of us who participate in Brookings Panel events) that obsesses over the central bank. The authors show that small firm managers resemble households in New Zealand much more than they do professional forecasters. I am sure that is true in the United States and elsewhere, too.

Almost all business managers and households have more important things to do—in their own view—than think about the Reserve Bank of New Zealand (RBNZ) or the Federal Reserve. This relative ignorance about monetary policy could be a case of *rational inattention* because the covariance between the firm's optimal relative price and the overall price level is small, as Christopher Sims (2003) discusses. Or it could be a

case of *irrational inattention*, that is, of people simply caring more about "OPP" (other people's puppies) than about the macroeconomic environment in which they are doing business. For current purposes, it does not much matter which explanation dominates—though I am inclined toward the latter.

Looking for a shred of good news about the importance of inflationary expectations, the authors report that 35 percent of their respondents say they would be "more likely to increase prices" if they "thought overall prices in the economy over the next 12 months were going to rise by more than what you are currently forecasting." But is 35 percent good news? Does it mean overall inflation is irrelevant to 65 percent of firms?

The paper's second big message is that inflationary expectations vary a lot, both cross-sectionally and over time. In both of these respects, business managers differ dramatically from professional forecasters, who *do* dote on the central bank. (See, for example, the authors' dramatic figure 8.)

One particular aspect of this disjuncture bears emphasis: In this sample, as well as in other evidence, ordinary citizens systematically *over*estimate inflation. It is not at all surprising that inflationary expectations are wide of the mark, nor that they vary hugely across individuals. But why people's expectations are systematically too *high,* virtually always, is a bit of a mystery—and a good question for subsequent research. There is a terribly important lesson here, for all of us in this profession and, I would guess, for all readers of the *Brookings Papers*: Do not generalize from yourself; you are not the representative agent.

The authors' third lesson is that ordinary people's expectations of inflation are *not* anchored, even though experts' expectations are. And this is so even though the RBNZ apparently enjoys great credibility. This finding is a head-scratcher until one looks at the inflation history of New Zealand during the inflation-targeting era. After hovering near 2 percent early in the period, inflation fell to roughly 1 percent, then zoomed up to 4 percent, and thereafter bounced around in the 1.5–5 percent range before recently dropping to about 0. That behavior does not look very "anchored" to me. So why should people's expectations be anchored?

Anchoring, of course, brings up the topic of inflation targeting as a way to conduct monetary policy. To my mind, there are three main arguments for adopting inflation targeting. First, it constrains discretion, as many, including Ben Bernanke and others (1999), argued years ago. I guess that is true, although Mervyn King, as governor of the Bank of England, wrote many letters explaining why the central bank's discretion should not be so constrained. Second, inflation targeting allegedly enhances transparency.

Yes, posting a numerical inflation target is a step forward. But over the years I have noticed that many central banks talk like "inflation nutters" while acting like they have a dual mandate. Is that being transparent?

The third argument for inflation targeting is the focus of this paper: It should anchor inflationary expectations. But anchored expectations are actually a mixed blessing. It is clearly a good thing when the central bank wants to adopt highly expansionary monetary policies. But, as the authors point out, it may not be such a good thing when the central bank is trying to reduce the ex ante real interest rate by raising $E(\pi)$.

In any case, the central message of this paper is that inflation targeting in New Zealand has anchored some people's expectations (mainly experts') but not others' (almost the entire population). Which brings us back to the meaning of the variable $E(\pi)$ in macro models. Whose expectations are these? The answer, as hardly any macro models recognize, is this: They are many people's—and they differ.

So, for example, the expectations of monetary policy experts—a group that includes professional forecasters, bond traders, and a few economists—profoundly influence nominal and real interest rates. This may be the only place in the paper where I disagree with the authors. They write in their concluding section:

> The lack of awareness by the general public, including firm managers, of the actions and objectives of monetary policymakers in a low-inflation environment is particularly problematic in periods when central bankers are seeking to affect inflation expectations through forward guidance.

Well, no, it is not so problematic. It is bond traders who make bond prices, and flattening the yield curve is the usual goal of forward guidance. These folks, the abnormal minority, pay rapt attention to every word uttered by the central bank.

The inattention problem lies elsewhere. For example, the paper shows that (mostly small) business managers—who set their nominal, and hence relative, prices—pay scant attention to central bank policy. One may safely assume, I think, that workers bargaining over nominal, and hence real, wages dote even less on the central bank. Finally, consumers, whose spending, macro models assume, depends on the ex ante real interest rate, probably know more about puppies than about the central bank's inflation target—or about monetary policy in general. As indicated earlier, this inattention could be rational or irrational. Either way, it is inattention.

I close with a heresy. Maybe inflationary expectations are not quite as important as modern macroeconomics makes them out to be. I do not mean

to imply that expected inflation is irrelevant; that is way too heretical and almost certainly wrong. But maybe, for example, *lagged* inflation is as important—in a Phillips curve, say—as *expected* inflation.[1] At least lagged inflation is, for one period, anchored.

REFERENCES FOR THE BLINDER COMMENT

Bernanke, Ben S., Thomas Laubach, Frederic S. Mishkin, and Adam S. Posen. 1999. *Inflation Targeting: Lessons from the International Experience.* Princeton University Press.

Roberts, John M. 2005. "How Well Does the New Keynesian Sticky-Price Model Fit the Data?" *B.E. Journal of Macroeconomics: Contributions to Macroeconomics* 5, no. 1: Article 10.

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50, no. 3: 665–90.

COMMENT BY

LARS E. O. SVENSSON This paper by Saten Kumar, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko discusses the results of a recent (2013–15) survey about inflation expectations and knowledge of monetary policy among managers of New Zealand firms. The main results are that the average inflation forecasts are higher than both actual inflation and the inflation target for both short and long horizons; that the average perception of recent inflation is higher than actual inflation; that there are large disagreements about forecasts and recent inflation; that the firms express more uncertainty than professional forecasters; that the firms have little knowledge about monetary policy; and that along these metrics the firms are more similar to households than to professional forecasters. The authors summarize their main conclusion in the title of the paper, "Inflation Targeting Does Not Anchor Inflation Expectations," and in the statement "Our results are not favorable to policymakers."

My first comment is, "Compared with what?" In order to draw these conclusions from the authors' survey, one would like to have not an essentially one-time survey but a time series of survey results, ideally over a sample period including years both before and after inflation

1. For evidence, see Roberts (2005).

targeting was introduced in New Zealand. One would also like to compare results with other related surveys in New Zealand and with similar surveys in other economies that use inflation targeting. Only then could one more firmly judge whether inflation targeting stabilizes and anchors inflation expectations.

Second, I miss some policy conclusions. Suppose the results are true. Should the authorities, in particular the Reserve Bank of New Zealand (RBNZ), do something about them, and if so, what?

Regarding my first comment, as the authors note, in New Zealand there is another related survey, namely the RBNZ Survey of Expectations, which surveys a sample of economists, businesses, and industry leaders.[1] It is quarterly and starts in 1987. The authors make light of this survey, stating that its sample is very small, that the firms involved are typically very large, and that the sample is not random but largely convenience-based.

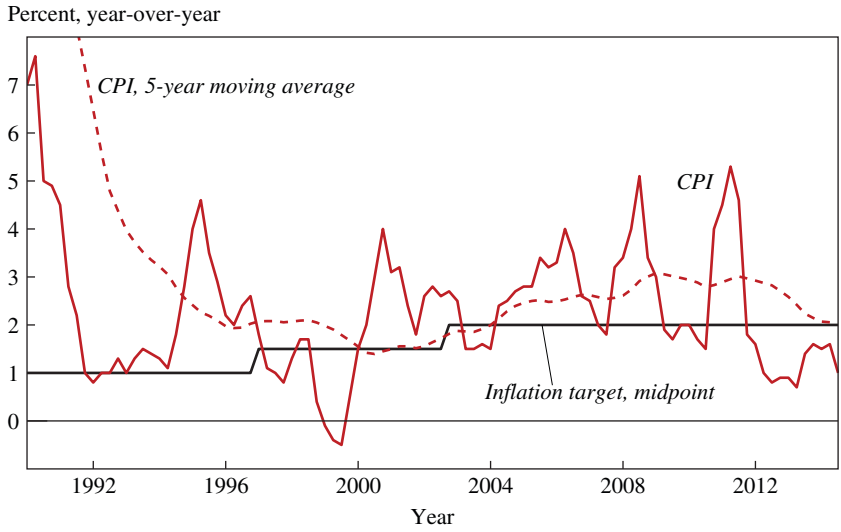
My figure 1 shows the annual CPI inflation rate in New Zealand, a 5-year (trailing) moving average of the inflation rate, and the midpoint of the target range. The target range was 0 to 2 percent from the beginning, changed to 0 to 3 percent in December 1996, and changed again to 1 to 3 percent in September 2002, shifting the target midpoint accordingly. We see that the inflation rate has fluctuated quite a bit, but that from the late 1990s the 5-year moving average has been either close to or somewhat above the target midpoint.

My figure 2 in addition shows the results of the RBNZ Survey of Expectations of the annual inflation rate 1 and 2 years ahead, respectively, with corresponding 5-year moving averages. We see that the inflation expectations are clearly influenced by the current inflation rate but vary less, and the 2-years-ahead inflation expectations are more stable than 1-year-ahead expectations.

In particular, the 5-year moving averages of inflation expectations are close to the 5-year moving averages of actual inflation. The respondents in the RBNZ survey seem to have, on average, fairly unbiased and therefore fairly realistic inflation expectations. The 5-year moving averages of the inflation expectations thus exceed the midpoint of the inflation target range as much as the 5-year moving average of the actual inflation rate does, rather than being anchored on the midpoint of the target range.

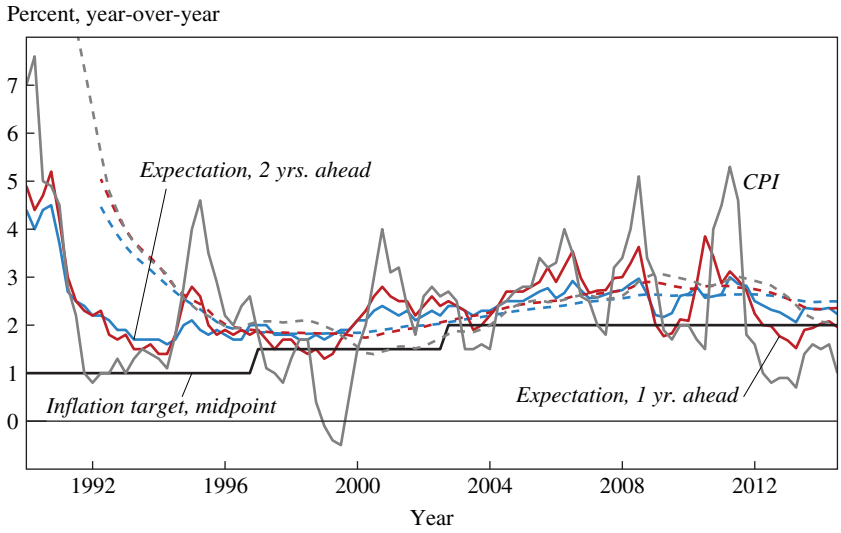
1. Information about the RBNZ's "M14 Survey of Expectations" can be found at http://www.rbnz.govt.nz/statistics/m14.

Figure 1. Inflation and Inflation Target, New Zealand, 1990–2014



Source: Datastream.

Figure 2. Inflation, Inflation Expectations, and Inflation Target, New Zealand, 1990–2014[a]



Sources: Datastream, RBNZ Survey of Expectations.
a. For each series, the dotted line represents a 5-year moving average.

It would have been desirable to have some comparison of the authors' surveys and the RBNZ surveys and a discussion of why they give such different impressions.

In Sweden, where the central bank, the Riksbank, announced in 1993 that an inflation target of 2 percent for the annual CPI inflation rate would apply from 1995 onward, there are several surveys of inflation expectations. One survey is the so-called Prospera Survey, commissioned by the Riksbank and conducted by TNS Sifo Prospera, which surveys the expectations of the annual CPI inflation rate 1, 2, and 5 years ahead among a panel of labor market organizations (trade unions and employers' associations), purchase managers, and money-market participants.[2] It began in 1995, has been done quarterly from 1996, and has been done monthly for money-market participants from 2009.

My figure 3 shows Sweden's annual CPI inflation rate, its 5-year moving average, and its average from 1995 up to each date. One can see that the CPI inflation rate has on average fallen substantially below the inflation target; in particular, the average inflation rate during the period 1995–2014 is only 1.2 percent, a full 0.8 percentage point below the target. The figure also shows the Prospera inflation expectations 1 and 2 years ahead and their corresponding 5-year moving averages.[3]

One can see that the inflation expectations are influenced by the current inflation rate and that the 2-years-ahead expectations are more stable than the 1-year-ahead ones. However, in contrast to the RBNZ Survey of Expectations, the moving averages are close to the inflation target rather than the moving average of actual inflation. Thus, the Prospera inflation expectations seem relatively strongly anchored on the inflation target, in spite of actual inflation falling substantially below the target.

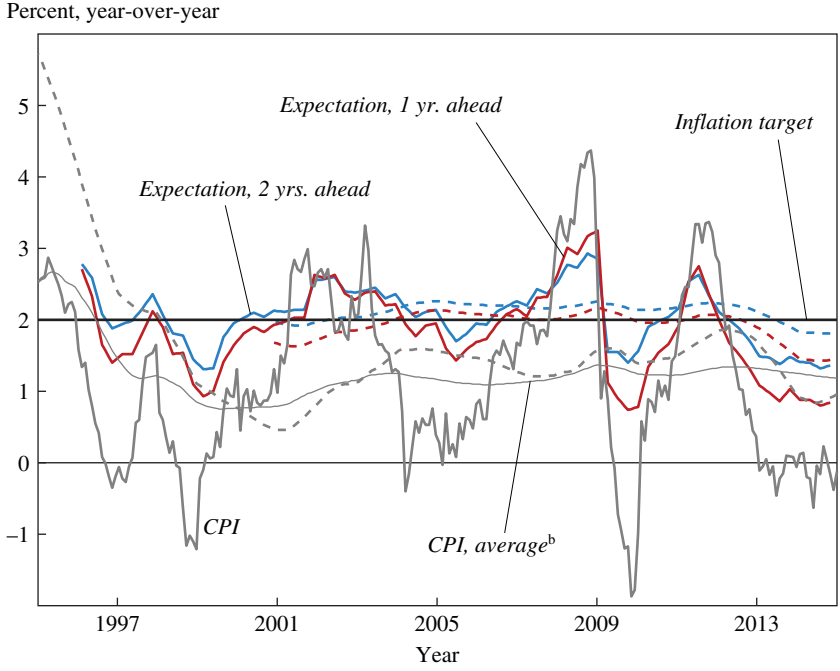
In Sweden there are two other relevant surveys, the Business Tendency Survey and the Consumer Tendency Survey, both conducted by the National Institute of Economic Research (NIER), a public authority under the Swedish Ministry of Finance.[4] The Business Tendency Survey

2. Information about the TNS Sifo Prospera's "Inflation Expectations" survey can be found at http://www.prospera.se/inflation-expectations.

3. In my previous work (Svensson 2011, 2015b), I discuss the reasons for and consequences of the systematic undershooting of the inflation target; Svensson (2015a) includes a comparison of the monetary policies of the RBNZ and the Riksbank.

4. Information about the NIER's "Economic Tendency Survey" can be found at http://konj.se/english/publications/economic-tendency-survey.html.

Figure 3. Inflation, Inflation Expectations, and Inflation Target, Sweden, 1995–2014[a]

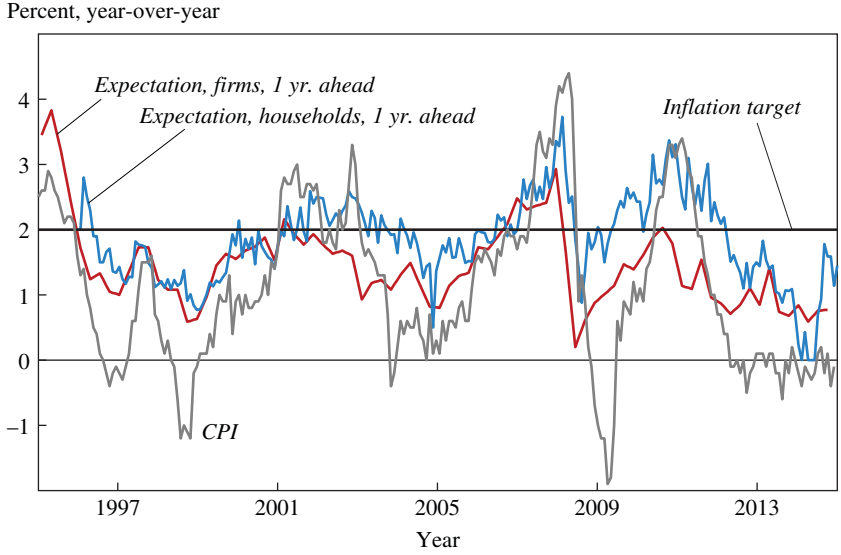


Sources: Statistics Sweden, TNS Sifo Prospera.
a. For each series, the dotted line represents a 5-year moving average.
b. Cumulative average of the CPI from 1995 onward.

is a large survey of firms in a set of relevant industries, with a sample of 6,500 firms, making up about 75 percent of employment in the total population of firms with activities in the relevant industries (including 100 percent of firms with 100 or more employees). The response rate is between 50 and 70 percent, depending on the industry. The survey is quarterly and started in 1987. The Consumer Tendency Survey uses a sample of about 1,500 households, is monthly, and started in 2002. Both surveys ask a number of different questions, including the respondent's expectation of the annual CPI inflation rate 1 year ahead.

My figure 4 shows the actual CPI inflation rate and the NIER surveys of firms' and households' expectations of inflation 1 year ahead. One can see that households' inflation expectations are on average close to the inflation target and above actual inflation. In contrast, firms' inflation expectations are on average below the target and close to average actual inflation. Thus, both Prospera and households' inflation expectations are more anchored on

Figure 4. Inflation, Inflation Expectations, and Inflation Target, Sweden, 1995–2015



Sources: Statistics Sweden, National Institute of Economic Research.

the inflation target and consequently biased and not rational. Firms' inflation expectations are not anchored on the inflation target but are more unbiased and rational.[5]

For the United States, the authors refer to the Michigan Survey of Consumers and the New York Fed Survey of Consumer Expectations. They argue that they find all the same patterns in inflation expectations as they

5. As I discussed in earlier work (Svensson 2015b), there are at least three observations that together indicate that inflation expectations in line with the target are more important than the NIER firms' inflation expectations in affecting wage setting in Sweden: (i) statements from the Swedish Trade Union Confederation and the Industrial Trade Unions, (ii) the fact that the TNS Sifo Prospera Survey reports inflation expectations of labor market organizations (both for employees and employers) similar to the expectations of all interviewees reported in my figure 2 (and thus close to the inflation target), and (iii) the importance of central wage negotiations over wage drift for wage setting after the introduction of the Industrial Cooperation and Negotiation Agreement in 1997. As I further discussed in Svensson (2015b), when nominal wages are negotiated and set under the expectation of an inflation rate equal to the 2 percent target, in spite of the average inflation rate falling significantly below 2 percent, the result is higher real wages than anticipated. This in turn leads to higher average unemployment than if inflation had on average been equal to the target. The average excess unemployment rate is estimated to be as large as 0.8 percentage point during 1997–2011, with a 95 percent confidence interval from 0.55 to 1.5 percentage points.

previously documented for managers of firms (as well as households) in New Zealand and conclude that expectations in the United States, therefore, appear just as unanchored as they do in New Zealand.

However, Michael Bryan, Brent Meyer, and Nicholas Parker (2015a) report results from the Federal Reserve Bank of Atlanta's Business Inflation Expectations Survey, a large monthly survey of businesses compiled by the Federal Reserve Bank of Atlanta since October 2011. Among other things, they document that, in the aggregate, firms' inflation expectations are very similar to the predictions of professional forecasters for national inflation statistics, despite a somewhat greater heterogeneity of expectations that they attribute to the idiosyncratic cost structure firms face. Bryan, Meyer, and Parker (2015a) also show that firms' inflation expectations bear little in common with the "prices in general" expectations reported by households. They additionally show that, during their 3-year sample, firms' inflation expectations appear to be unbiased predictors of their year-ahead observed (perceived) inflation. In a blog post shortly after Kumar, Afrouzi, Coibion, and Gorodnichenko presented their paper at Brookings, Bryan, Meyer, and Parker (2015b) suggest that their own research indicates that the authors' results are due to poorly phrased questions and that there is strong evidence that their respondents either did not understand the questions about "prices in general" or were misinterpreting them, as compared to questions about "inflation."

In summary, other surveys in New Zealand, Sweden, and the United States indicate better anchoring of inflation expectations on the inflation target or on average actual inflation. The reasons for these discrepancies are not well understood, and the precise formulation of survey questions appears to matter. The authors' strong conclusions from a one-off survey, including any policy conclusions, therefore seem premature.

REFERENCES FOR THE SVENSSON COMMENT

Bryan, Michael F., Brent H. Meyer, and Nicholas B. Parker. 2015a. "The Inflation Expectations of Firms: What Do They Look Like, Are They Accurate, and Do They Matter?" Working Paper no. 2014-27a. Federal Reserve Bank of Atlanta.

———. 2015b. "What Do U.S. Businesses Know That New Zealand Businesses Don't? A Lot (Apparently)." *macroblog,* September 21. Federal Reserve Bank of Atlanta.

Svensson, Lars E. O. 2011. "Practical Monetary Policy: Examples from Sweden and the United States." *Brookings Papers on Economic Activity,* Fall: 289–332.

———. 2015a. "Forward Guidance." *International Journal of Central Banking* 11, suppl. 1: 19–64.

———. 2015b. "The Possible Unemployment Cost of Average Inflation below a Credible Target." *American Economic Journal: Macroeconomics* 7, no. 1: 258–96.

GENERAL DISCUSSION William Brainard opened the discussion by remarking that he enjoyed the paper by Saten Kumar, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko in part because it confirmed much of what he had been persuaded of by earlier work. He mentioned Truman Bewley, who sampled more than 500 firms and found that none of them indicated ever paying attention to the Federal Reserve's targeting in determining their own pricing. Brainard had heard the same in talking with businessmen. However, it did strike him as surprising that this paper included a significant number of financial services firms, which he had always thought paid great attention to what was going to happen to the bond prices and the stock market and, one would assume, therefore listened to what the Federal Reserve was announcing. Did the authors find that knowledge of the Reserve Bank of New Zealand behavior was stronger at least in that financial services subsample?

Brainard also wondered whether the authors' survey also asked firms about their price setting and inflation expectations specifically regarding the cost of the materials that they buy. After all, most of the firms that set actual prices that later show up in the CPI are producing only a tiny part of it, and many others produce intermediate products that are not averaged into the CPI at all. Overall, though, it was not surprising to him that many people do not think about the connections between their personal shopping experience, their own firm's pricing, and a third thing that they are not directly involved in, which is the bundle of goods that make up the CPI.

Ben Friedman said he liked the paper for the same reasons discussant Alan Blinder outlined. He thought it was a refreshing antidote to the usual narcissism of people in the economics profession who do not recognize that everybody else finds reading Federal Open Market Committee statements much less interesting than reading the sports pages or crime thrillers. He thought the key question, which discussant Lars Svensson also raised, was this: What is the policy implication? The paper did not say much about that, yet in the United States it is clearly very important, following the paper's own argument.

Here in the United States, Friedman reminded everyone, the debate over what the inflation target ought to be remains unsettled. As one argument runs, if 2 percent seemed right 10 years ago, the experience of the post-crisis period would surely indicate the optimal number should be higher. A standard rebuttal to that argument has been that one would of course set a higher number if we were redesigning the system from scratch, but because the public is so focused on the 2 percent number already promoted, raising it—even to 2.5 percent—would undermine the Federal Reserve System's credibility. He felt the evidence from this paper refutes that line of thinking, since it shows that only a small segment of the public pays attention to the Federal Reserve at all. Clearly, a large part of the population would not be bothered one way or the other by a change in the inflation target. The paper's finding, he said, is therefore a valuable contribution to policy thinking.

Having attended the NBER Summer Institute the previous summer, where Governor Kuroda from the Bank of Japan spoke, Martin Feldstein reported that Kuroda had been pleased with household surveys showing that their inflation expectation centered around 2 percent. To Kuroda this indicated his policies were working, notwithstanding that the distribution of responses was quite flat, ranging all the way from −10 percent to +15 percent. Feldstein recalled a Michigan survey some years earlier that also found a very wide dispersion of expected inflation rates. And yet, the Federal Reserve always reports that inflation expectations are well anchored. Although he understood Friedman's point about the public not knowing whether the Federal Reserve was targeting 3 percent instead of 2 percent, Feldstein believed that whenever it raised its inflation target there would still be a headline that people would notice, and although they might not understand what it meant, it would be perceived as an event. In light of all that, he wondered what economists should think about the fact that mean expectations are so close to the target, and that at the same time the distribution is so very wide. What are the implications of that for policy and also for the dynamics of the expectations process itself?

Annamaria Lusardi echoed Blinder's point that the paper was very consistent with the household survey, though she also felt the word "manager" should not mislead one to think that the authors' survey of businesses tapped a very different population from households, since it seemed to survey small firms. And she noted that there is a lot of evidence that the population as a whole is financially illiterate when it comes to inflation and the consequences of inflation, and she was somewhat disappointed that the paper did not look into the effects of that ignorance.

In this regard, she disagreed with Blinder that the implication was that it does not matter if people do not understand inflation, since that has implications for how people run businesses, how they set prices and wages, and how they borrow in the market. If other waves of surveys can be run to document these behaviors it would enable us to know whether inflation expectations really do matter, or not.

She appreciated the comparisons with the United States' experience, but noted that New Zealand was the pioneer in targeting inflation, starting in 1990. If after 25 years people there still have not adjusted to that policy, this causes one to wonder what has been happening in other countries that have chosen an inflation target. She seemed to recall that in his research, Frederic Mishkin found that New Zealand had a policy of firing the central banker if he did not meet the targets, and if that is correct then it implies tremendous weight being placed on this approach.

The U.S. surveys she has reviewed show that some learning about inflation has taken place over the years, but generally the learning occurs due to inflationary episodes. Perhaps monetary policy has been communicated ineffectively—as something so boring people cannot even remember it—and if so then maybe this is what needs to be improved. If that is the policy implication of this innovative and provocative paper, then it suggests a straightforward solution, although how to communicate effectively remains an open question.

Carmen Reinhart wondered if her colleagues in the discussion were not overinterpreting the paper's results. She pointed out that in Buenos Aires today, the cab drivers are talking about inflation and how INDC, the government statistical agency, has been mismeasuring it, and they also talk about the exchange rate and about parallel markets. In other words, the average person there has a pretty well-informed view, precisely because inflation is a big problem. The paper happens to be comparing the United States and New Zealand at a time when inflation is not affecting households' bottom line. By contrast, if one looks further back, before the 1990s, when inflation was in double digits both in New Zealand and in the United States, cost-of-living adjustments were a major concern. Reinhart wondered if it would be possible to examine these other periods and see whether the paper's results would have been significantly different, even though those are periods before inflation targeting. It seemed to her quite possible that inflation targeting does anchor expectations during times when inflation is biting into people's lives.

Robert Gordon felt the paper raised very important questions about the concepts of inflation targeting, anchored expectations, and forward

guidance. While concurring with Reinhart's observation about the high level of popular understanding of inflation and central bank policies in high-inflation countries, he pointed out that the paper is actually focused on low-inflation countries. In his view, the role of expectations here has been greatly exaggerated. There is an alternative framework that links current inflation expectations to past inflation, best summarized by John Taylor's 1980 model of overlapping contracts. He also pointed to Olivier Blanchard's input/output model of inflation, presented at a 1987 Brookings Panel, which emphasized the role of intermediate goods' prices. This in turn led him to recall what Truman Bewley had found when investigating how businesses set individual prices: Business managers did not consider either the Federal Reserve's targets or aggregate reported prices but, simply, their own costs, which in many cases consisted more of intermediate goods purchases than wages.

Businesses look at their profits and at competitors' prices, Gordon reminded everyone, and when there is a downturn in the macro economy they first notice that their sales have fallen off and, consequently, put things on sale. If they see their inventories building up, they respond, so there is a basic demand effect. All of this is in effect a giant backward-looking mechanism, not one based on expectations. Businesses' pricing decisions depend on previous price decisions by intermediate-goods suppliers, who in turn are dependent on even earlier intermediate-goods price decisions. In short, he argued, in low-inflation countries, inflation expectations play very little role.

Christopher Carroll attempted to synthesize what had been said by various colleagues in the discussion to that point. People in New Zealand do not have clear expectations about inflation, but neither does it follow that Friedman was correct in his suggestion that increasing the inflation target would have little effect, since the action would still create headlines the next day. This suggested to Carroll that what is needed is a theoretical framework that links people's expectations of inflation with the environment they are living in, such as the number of news stories published on inflation, something the University of Michigan's Consumer Sentiment Index found had an impact. Carroll mentioned a paper he wrote years earlier in which he had found that household expectations are not rational in the traditional sense. Households did move in the direction of the expectations of the professional forecasters, he had found, but they did so gradually, updating their expectations by roughly 25 percent of what the professionals believed and being more apt to do so when there were more news stories. In today's discussion, he found himself agreeing with a point

raised by both Reinhart and Gordon, namely that it matters a lot whether what people believe about inflation is forward looking or backward looking. That is a deeply important question for monetary policy. All of this suggested to him that the research agenda going forward should aim to figure out, once the inflation rate has gone up, when it is that people start paying more attention. Is it the news stories, which are future-oriented, that trigger their attention, or is it noticing what has happened in the past to prices? Presumably macro models will assign different implications to each of those behaviors.

Valerie Ramey reminded the others that expectations only matter if they translate into action. She speculated that faulty expectations were not necessarily distorting the pricing decisions of small firms. Echoing Gordon's comments, she believed that while large firms pay close attention to the general inflation level and adjust their expectations to it, small firms care only about what their competitors are charging and what their input costs are; the general level of inflation does not concern them. In support of Reinhart's examples from Argentina, Ramey mentioned recent research that found that individuals update their inflation expectations more quickly in high inflation countries, such as Argentina, than in low inflation countries, such as the United States. Ramey speculated that people do not keep track of inflation when it is low. If that is so, then policies such as Abenomics may be ineffective.

Ricardo Reis commented that the title of the paper was slightly misleading, because whether inflation targeting anchors expectations is ultimately a relative question—relative to what other policy regime is it anchoring them? He has seen three or four cross-national studies of inflation targeting that have looked at the response from professionals and households, and all found that inflation targeting does anchor inflation expectations in the sense of lowering average forecast errors and narrowing the dispersion of expectations. It does not drive dispersion down to zero, and there can be a lot of dispersion remaining, as in New Zealand, but it does impart an effect relative to the time series.

The other quarrel he had with the authors' findings is that they lacked a comparative context. How did the survey responses on inflation compare with expectations of other economic measures? Reis conjectured that one would find much larger errors in people's knowledge of nominal GDP growth, both current and forecasted, and likewise in knowledge of the unemployment rate, despite the fact that unemployment matters quite a bit for elections. The same, he conjectured, would be true for knowledge of the Federal Reserve's current and near-term interest rates. Looked at

in this relative context, the central bank's announcement of a target for this macro variable—targeted inflation—may actually be showing some effect, in comparison with nominal GDP or unemployment, which are not announced as targets.

One could also look at these data, Reis suggested, and think of the many models of expectations that are out there, including epidemiological models, inattention models, and various learning models. But it was unclear to him that the data reported in this paper either rejected any of those models or was consistent with them because of a problem familiar in applied microeconomics: To find clear results, one needs a shock to identify whether the effect is consistent with the model or not. At the same time, Reis argued for the use of surveys to pin down the right models, rather than abandoning the writing of models altogether just because the link between policy regimes and expectations is unclear. Policy actions like inflation targeting do affect expectations and those expectations, as Ramey noted, affect outcomes. This feedback loop between expectations and actions, which goes in both directions, is what we need both data and models to better design.

Donald Kohn wanted to underline a final point Svensson had raised in his discussant remarks, that it is the time series that matters. More than the mean or the median of misperceived inflation, the flatness of the distribution, or how frequently people are revising data, what matters to policymakers is whether changes in inflation expectations are reinforcing boom and bust cycles, as when people spend more now to avoid higher prices later or when they hold back because they fear deflation. In Kohn's view, the change in expectations is what is most important. In the Michigan household survey the longer-term inflation expectations were remarkably stable throughout a very turbulent period in the economy, and that was helpful to the economy as a whole. It can be contrasted with what happened in the 1970s, when a large rise in inflation caused people to revise their expectations even higher, leading to a bad cycle.

Kumar responded to the discussion, first by following up on Reis's point that business managers are not behaving randomly but trying to earn a profit and, therefore, digesting information as efficiently as possible. He mentioned an earlier paper of his own that employed a quasi-randomized experiment, in which some people were given additional information in order to measure how they would incorporate it into their expectations and decisions. The behavior they observed turned out to be perfectly rational, so one can conclude that the problem with inflation expectations stems from information constraints rather than from irrational thinking or stupidity.

As far as the time series question, Kumar said he and the coauthors would have liked to have had access to long time series, but they were not available. Further cross-sectional analysis, though, is not only desirable but possible, and since their research team included an economist from New Zealand, another from France, another from Ukraine, and another from Iran, they are already trying to generate this survey in these countries along with the United States.

In response to the comments that people in the United States do not seem to care about inflation, he countered that this is plainly untrue. In polls there, 80 percent of people say they are hurt by inflation, and when asked to identify priorities for the president of the United States most people cite inflation as much more important even than the war in Iraq.

Kumar noted that research comparing communication about inflation in Sweden with that in the United States was very telling. In Sweden the survey asks a four-word question and allows a four-word answer. In the United States, even at a press conference the question is one paragraph long, followed by four or five paragraphs of answers with a lot of ifs and whens. Another study found that in order to understand the minutes of the Federal Reserve Board meetings, back in the early 1990s a high school diploma was sufficient, but today it requires a Ph.D. The public simply cannot understand the purpose of monetary policy with this kind of communication.

He agreed that inflation targeting may be very instrumental in narrowing the range of possibilities people will consider when they think about their inflation expectations, but at the same time, in a low-inflation environment it is very hard to push this effect far. Considering the wide distribution of beliefs their study found in New Zealand, Kumar concluded that the outcome is certainly very different from what the standard Keynesian models would predict, which would be everyone in agreement based on rational expectations. One thing is certain: The world is not there.

MICHAEL BOLDIN
Federal Reserve Bank of Philadelphia

JONATHAN H. WRIGHT
Johns Hopkins University

Weather-Adjusting Economic Data

ABSTRACT This paper proposes and implements a statistical methodology for adjusting employment data for the effects of deviations in weather from seasonal norms. This is distinct from seasonal adjustment, which controls only for the normal variation in weather across the year. We simultaneously control for both of these effects by integrating a weather adjustment step in the seasonal adjustment process. We use several indicators of weather, including temperature and snowfall. We find that weather effects can be important, shifting the monthly payroll change number by more than 100,000 in either direction. The effects are largest in the winter and early spring months and in the construction sector. A similar methodology is constructed and applied to data in the national income and product accounts (NIPA), although the manner in which NIPA data are reported makes it impossible to integrate weather and seasonal adjustments fully.

Macroeconomic time series are affected by the weather. In the first quarter of 2014, real GDP contracted by 0.9 percent at an annualized rate. Commentators and Federal Reserve officials attributed part of the decline to an unusually cold winter and large snowstorms that hit the East Coast and the South during the quarter (Macroeconomic Advisers 2014; Yellen 2014).[1] Similarly, the slowdown in growth in the first quarter of 2015 was widely ascribed to another exceptionally harsh winter and other transitory factors (Yellen 2015). While the effects of regular variation in

1. In November 2013, the Survey of Professional Forecasters expected a seasonally adjusted increase of 2.5 percent in 2014Q1. The original report for the quarter was 0.1 percent, later revised to –2.1 percent, and subsequently revised to –0.9 in the 2015 annual NIPA adjustments that included revisions to the seasonal adjustment process, as discussed in section III below. With a snapback rate of 4.6 percent in the second quarter, it is highly plausible that weather played a significant role in the decline.

weather within a year should, in principle, be taken care of by the seasonal adjustment procedures that are typically applied to economic data, these adjustments are explicitly not supposed to adjust for variations that are driven by deviations from the weather norms for a particular time of year. It is typically cold in February, depressing activity in some sectors, and seasonal adjustment controls for this. But seasonal adjustment does not control for whether a particular February is colder or milder than normal.

Our objective in this paper is to construct and implement a methodology for estimating how the data would have appeared if weather patterns had followed their seasonal norms. Monetary policymakers view weather effects as transitory—given the long and variable lags in monetary policy, policymakers do not generally seek to respond to weather-related factors. It follows from this that the economic indicators they are provided with ought, as far as possible, to be purged of weather effects. Moreover, we argue that failing to control for abnormal weather effects distorts conventional seasonal adjustment procedures.

The measurement of inflation provides a useful analogy. The Federal Reserve focuses on core inflation, excluding food and energy, rather than headline inflation. The motivation is not that food and energy are inherently less important expenditures but that fluctuations in their inflation rates are transitory. Core inflation is more persistent and forecastable, and indeed a forecast of core inflation may be the best way of predicting overall inflation (Faust and Wright 2013). In the same way, economic fluctuations caused by the weather are real, but they are transitory. We may obtain a better measure of the economy's underlying momentum by removing the effects of abnormal weather.

Economists have studied the effects of the weather on agricultural output for a long time, going back to the work of R. A. Fisher (1925). More recently, they have also used weather as an instrumental variable (see, for example, Miguel, Satyanath, and Serengeti [2004]), arguing that weather can be thought of as an exogenous driver of economic activity. Statistical agencies sometimes judgmentally adjust extreme observations due to specific weather events before applying their seasonal adjustment procedures.[2] Although there is a long literature on seasonal adjustment, we are aware of only a few papers on estimating the effect of unseasonal weather on

2. Even when agencies do this, their goal is just to prevent the anomalous weather from distorting seasonals, not to actually adjust the data for the effects of the weather. We discuss this in more detail later.

macroeconomic aggregates. The few papers on the topic include those by Macroeconomic Advisers (2014), which regresses seasonally adjusted aggregate GDP on snowfall totals, estimating that snow reduced 2014Q1 GDP by 1.4 percentage points at an annualized rate; by Justin Bloesch and François Gourio (2014), who likewise study the relationship between weather and seasonally adjusted data; by Melissa Dell, Benjamin Jones, and Benjamin Olken (2012), who implement a cross-country study of the effects of annual temperature on annual GDP; and by Christopher Foote (2015), who studies weather effects on state-level employment data. None of these papers integrates weather adjustment into the seasonal adjustment process, however. This is what the current paper attempts to do.

We focus mainly, but not exclusively, on the seasonal adjustment of the Bureau of Labor Statistics (BLS) Current Employment Statistics (CES) survey (the "establishment" survey), which includes total nonfarm payrolls. We do so because it is clearly the most widely followed monthly economic indicator, and also because it is an indicator for which researchers can approximately replicate the official seasonal adjustment process, unlike the NIPA data. We consider simultaneously adjusting these data for both seasonal effects and unseasonal weather effects. This can be quite different from ordinary seasonal adjustment, especially during the winter and early spring. Month-over-month changes in nonfarm payrolls are in several cases higher or lower by as much as 100,000 jobs when using the proposed seasonal-and-weather adjustment rather than ordinary seasonal adjustment. Using seasonal-and-weather adjustment increases the estimated pace of employment growth in the winters of 2013–14 and 2014–15.

The plan for the remainder of this paper is as follows. In section I, we discuss alternative measures of unusual weather and evaluate how they relate to aggregate employment. This is intended to give us guidance on which weather indicators have an important impact on employment data. In section II, we describe seasonal adjustment in the CES and discuss how adjustment for unusual weather effects may be added into this—seasonal adjustment is implemented at the disaggregate level. In section III we extend the analysis to NIPA data. Section IV concludes.

I. Measuring Unusual Weather and Its Effect on Aggregate Employment Data

We need to construct measures of unseasonal weather that are suitable for adjusting the CES survey. We first obtained data from the National Centers for Environmental Information on daily maximum temperatures,

precipitation, snowfall, and heating degree days (HDDs)[3] at one station in each of the largest 50 metropolitan statistical areas (MSAs) by population, in the United States from 1960 to the present. The stations were chosen to provide a long and complete history of data,[4] and are listed in table 1. We averaged these across the 50 MSAs, with the averages weighted by population, determined from the 2010 census. This was designed as a way of measuring U.S.-wide temperature, precipitation, and snowfall in a way that makes a long time series easily available and that puts the highest weight on areas with the greatest economic activity. Weather, of course, varies substantially around the country, and it might seem more natural to adjust state-level employment data for state-level weather effects. We used national-level employment data with national-level weather because the BLS produces state and national data separately using different methodologies. National CES numbers are quite different from the "sum of states" numbers, because both state and national CES numbers are constructed by survey methods, whereas the national data use more disaggregated cells. Meanwhile, it is the national numbers that garner virtually all the attention from Wall Street and the Federal Reserve.

Let $temp_s$ denote the actual average temperature on day s, and define the unusual temperature for the day as $temp_s^* = temp_s - \frac{1}{30}\sum_{y=1}^{30} temp_{s,y}$, where $temp_{s,y}$ denotes the temperature on the same day y years previously. Likewise, let prp_s^*, $snow_s^*$, and hdd_s^* denote the unusual precipitation, snowfall, or HDD on day s, relative to the 30-year average. This is in line with the meteorological convention of defining climate norms from 30-year averages (World Meteorological Organization 2011).

In assessing the effect of unusual weather on employment as measured in the CES, we want to take careful account of the within-month timing of the CES survey. The CES survey relates to the pay period that includes the 12th day of the month. Some employers use weekly pay periods, others use biweekly periods, and a few use monthly periods. A worker is counted if she works at any point in that pay period. Cold weather or snow seems

3. The HDD at a given station on a given day is defined as max $(18.3 - \tau, 0)$, where τ is the average of maximum and minimum temperatures in degrees Celsius.

4. An alternative measure of snowfall, used by Macroeconomic Advisers (2014), is based on a data set of daily county-level snowfall maintained by the National Centers for Environmental Information. This clearly has the advantage of greater cross-sectional granularity. However, these data only go back to 2005. Our data go much further back, allowing us to construct a longer history of snowfall effects and to measure normal snowfall from 30-year averages.

Table 1. Weather Stations Used to Measure National Weather[a]

MSA	*Station*	*MSA*	*Station*
New York	Central Park	San Antonio	San Antonio Intl. Airport
Los Angeles	Los Angeles Intl. Airport	Orlando	Orlando Intl. Airport
Chicago	Chicago O'Hare Intl. Airport	Cincinnati	Cincinnati/Northern Kentucky Intl. Airport
Dallas	Dallas/Fort Worth Intl. Airport	Cleveland	Cleveland Hopkins Intl. Airport
Philadelphia	Philadelphia Intl. Airport	Kansas City	Kansas City Intl. Airport
Houston	George Bush Intcntl. Airport	Las Vegas	McCarran Intl. Airport
Washington	Washington Dulles Intl. Airport	Columbus	Port Columbus Intl. Airport
Miami	Miami Intl. Airport	Indianapolis	Indianapolis Intl. Airport
Atlanta	Hartsfield-Jackson Intl. Airport	San Jose	Los Gatos
Boston	Logan Intl. Airport	Austin	Camp Mabry
San Francisco	San Francisco Intl. Airport	Virginia Beach	Norfolk Intl. Airport
Detroit	Coleman A. Young Intl. Airport	Nashville	Nashville Intl. Airport
Riverside	Riverside Fire Station	Providence	T. F. Green Airport
Phoenix	Phoenix Sky Harbor Intl. Airport	Milwaukee	Gen. Mitchell Intl. Airport
Seattle	Seattle-Tacoma Intl. Airport	Jacksonville	Jacksonville Intl. Airport
Minneapolis	Minneapolis-Saint Paul Intl. Airport	Memphis	Memphis Intl. Airport
San Diego	San Diego Intl. Airport	Oklahoma City	Will Rogers World Airport
St. Louis	Lambert-St. Louis Intl. Airport	Louisville	Louisville Intl. Airport
Tampa	Tampa Intl. Airport	Hartford	Bradley Intl. Airport
Baltimore	Baltimore/Washington Intl. Airport	Richmond	Richmond Airport
Denver	Stapleton/Denver Intl. Airport[b]	New Orleans	Louis Armstrong Intl. Airport
Pittsburgh	Pittsburgh Intl. Airport	Buffalo	Buffalo Niagara Intl. Airport
Portland (Ore.)	Portland Intl. Airport	Raleigh	Raleigh-Durham Intl. Airport
Charlotte	Charlotte Douglas Intl. Airport	Birmingham	Birmingham Airport
Sacramento	Sacramento Executive Airport	Salt Lake City	Salt Lake City Intl. Airport

a. This table lists the 50 weather stations used to construct national average daily temperature, snowfall, and HDD data. Each weather station corresponds to one of the 50 largest MSAs by population in the 2010 Census.
b. Stapleton International Airport was replaced by Denver International Airport in 1995.

most likely to affect employment status on the day of that unusual weather, but it is also possible that, for example, heavy snow might affect economic activity for several days after a snowstorm has ceased. Putting all this together, temperature/snowfall conditions in the days up to and including the 12th day of the month are likely to have some effect on measured employment for that month. The further before the 12th day of the month the unusual weather occurred, the less likely it is to have affected a worker's employment status in the pay period bracketing the 12th, and so the less important it should be. It is hard to know a priori how to weight unusual weather on different days up to and including the 12th day of the month, but, on the other hand, it seems likely that unusual weather after the 12th day of the month ought to have little effect on employment data for that month.[5]

In solving this problem, we try to let the data speak. Our proposed approach assumes that the relevant temperature/precipitation/snowfall conditions are a weighted average of the temperature/precipitation/snowfall in the 30 days up to and including the 12th day of the month, using a Mixed Data Sampling (MIDAS) polynomial as the weights to avoid overfitting. We want to use this specification to collapse the daily weather data that we have into monthly weather measures. We will spell out the details of the MIDAS polynomial and its estimation below. MIDAS polynomials were proposed by Eric Ghysels, Pedro Santa-Clara, and Rossen Valkanov (2004, 2005) and by Elena Andreou, Ghysels, and Andros Kourtellos (2010) as a device for handling mixed frequency data in a way that is parsimonious yet flexible—exactly the problem that we face here. The presumption is that unusual weather on or just before the 12th day of the month should get more weight than unusual weather well before this date.

In addition to temperature, precipitation, snowfall, and HDDs, there are two other weather indicators that we consider. First, as an alternative way of measuring snowfall, the National Centers for Environmental Information produce regional snowfall indexes that measure the disruptive impact of significant snowstorms. These indexes take into account the area affected by the storm and the population in that area, for six different regions of the

5. There are actually ways in which weather after the 12th could matter for CES employment that month. For example, suppose that a new hire was planning to begin work on the 13th and the 13th happens to be the last day of the pay period. She would be counted as employed in that month. But if bad weather caused the worker's start date to be delayed, then she would not be defined as employed in that month. However, we do evaluate the possibility that weather just after the 12th could affect employment for that month.

country. See Paul Kocin and Louis Uccellini (2004) and Michael Squires and others (2014) for a discussion of these regional snowfall impact (RSI) indexes. They are designed to measure the societal impacts of different storms, which make them potentially very useful for our purposes. They have the drawback that they do not cover the western part of the country, but there are only two big cities that are not covered and that receive significant snowfall: Denver and Salt Lake City.

Any snowstorm affecting a region has an index value, a start date, and an end date. We treat the level of snowfall in that region as being equal to the index value from the start to the end date, inclusive. For example, a storm affecting the southeast region was rated as 10.666, started on February 10, 2014, and ended on February 13, 2014. We treat this index as having a value of 10.666 on each day from February 10 to 13, 2014. For each of those days, we then create a weighted sum of the six regional snowstorm indexes to get a national value, where the weights are the populations in the regions (from the 2010 Census). We then used this RSI index as an alternative to the average snowfall. Second, the household Current Population Survey (CPS) asks respondents if they were unable to work because of the weather. We seasonally adjust the number who were absent from work[6] in month t, using the default X-13 filter, and then treat this variable, abs_t, as an additional weather indicator.

We first estimate eight candidate models giving the effects of different weather measures on aggregate employment. Intuitively, we are simply interested in regressing monthly aggregate not seasonally adjusted (NSA) employment onto a weighted average of daily weather data, where the weights give the best possible fit. This is intended as a precursor to incorporating weather effects in CES seasonal adjustment. However, weather is only a very small part of what drives aggregate employment. We also want the model to allow for trend and seasonal components.

I.A. Eight Candidate Models

Each of our eight candidate models is an "airline model"—the default model in the first stage of the X-13—fitted to aggregate NSA employment, but augmented by weather variables. Each model specifies that there are trend and seasonal components that are nonstationary and consequently require taking first differences and differences from the same month one

6. This is the number with a job, not at work, in nonagricultural industries (series LNU02036012).

year earlier. After this differencing, the employment data are driven by weather effects and by moving average errors. The specific model is of the form

$$(1) \qquad (1 - L)(1 - L^{12})(y_t - \gamma' x_t) = (1 + \theta L)(1 + \Theta L^{12})\varepsilon_t,$$

where y_t is total NSA employment for month t, L is the lag operator, and ε_t is an independent and identically distributed error term. The eight models differ only in the specification of the regressors in x_t. The specifications that we consider are as follows:

SPECIFICATION 1: TEMPERATURE ONLY There are 12 elements in x_t, each of which is $\sum_{j=0}^{30} w_j temp^*_{s-j}$ interacted with one of 12 monthly dummies, where day s is the 12th day of month t, and where

$$w_j = B\left(\frac{j}{30}, a, b\right)$$

and

$$B(x; a, b) = \frac{\exp(ax + bx^2)}{\sum_{j=0}^{30} \exp\left(a\frac{j}{30} + b\left(\frac{j}{30}\right)^2\right)}.$$

$B(x; a, b)$ is the MIDAS polynomial. In all, this model has 17 parameters: the 12 elements of γ along with a, b, θ, Θ, and the variance of the error term. Temperature is interacted with month dummies. The motivation for this is that the effect of temperature on the economy depends heavily on the time of year. For example, unusually cold weather in winter lowers building activity, but unusually cold weather in the summer might have little effect on this sector, or might even boost it. Likewise, warm weather boosts demand for electricity in summer but weakens demand for electricity in winter.

SPECIFICATION 2: HDD ONLY There are 12 elements in x_t, each of which is $\sum_{j=0}^{30} w_j hdd^*_{s-j}$ interacted with one of 12 monthly dummies, where $w_j = B\left(\frac{j}{30}, a, b\right)$.

SPECIFICATION 3: TEMPERATURE AND SNOWFALL There are 13 elements in x_t. The first 12 are as in specification 1. The 13th element is $\sum_{j=0}^{30} w_j snow^*_{s-j}$,

where $snow_s^*$ denotes the unusual snowfall on the 12th day of month t, measured as the population-weighted average across the 50 MSAs. The monthly snow variable is not interacted with month dummies, because it falls only in the winter months, and its effect on employment is likely to be similar in any winter month.

SPECIFICATION 4: TEMPERATURE AND SNOWFALL (RSI INDEX) The specification is as in specification 3, except using the RSI index to measure snowfall.

SPECIFICATION 5: TEMPERATURE, SNOWFALL (RSI INDEX), AND WEATHER-RELATED ABSENCES FROM WORK The specification is the same as in specification 4 except that abs_t is included in the 14th element of x_t.

SPECIFICATION 6: TEMPERATURE, SNOWFALL (RSI INDEX), AND PRECIPITATION There are 14 elements in x_t. The first 13 are as in specification 4. The 14th element is $\sum_{j=0}^{30} w_j prp_{s-j}^*$, where prp_s^* denotes the unusual precipitation on the 12th day of month t, measured as the population-weighted average across MSAs.

SPECIFICATION 7: TEMPERATURE, SNOWFALL (RSI INDEX), AND LAGS OF TEMPERATURE AND SNOWFALL There are 13 elements in x_t. Each of the first 12 is $\sum_{j=0}^{90} w_j temp_{s-j}^*$ interacted with one of 12 monthly dummies, where $w_j = B\left(\frac{j}{30}, a, b\right)$ for $j \le 30$, $w_j = c$ for $31 \le j \le 60$, and $w_j = d$ for $j > 60$. The last element is $\sum_{j=0}^{90} w_j snow_{s-j}^*$. In this specification, the parameters c and d determine the weight of weather two and three months prior.

SPECIFICATION 8: TEMPERATURE, SNOWFALL (RSI INDEX), AND TEMPERATURE AND SNOWFALL JUST AFTER THE CES SURVEY DATE There are 13 elements in x_t. Each of the first 12 is $\sum_{j=-2}^{90} w_j temp_{s-j}^*$ interacted with one of 12 monthly dummies, where $w_j = B\left(\frac{j}{30}, a, b\right)$ for $j \ge 0$ and $w_j = c$ otherwise. The last element is $\sum_{j=-2}^{90} w_j snow_{s+j}^*$. In this specification, we use a MIDAS-weighted average of the days up to and including the 12th, and an extra parameter c determines the weight of weather on the 13th and 14th of the month.

Note that in all these specifications, we are assuming that the effect is linear in weather; unusually cold and unusually warm temperatures are assumed to have effects of equal magnitude but opposite sign.

All the weather indicators that we consider are physical measures of weather that are essentially exogenous, except for self-reported work

absences due to weather (specification 5).[7] We are consequently a little more cautious about the use of weather-related work absences as a weather measure. Of course, it could be that this variable is giving us more information about the economic costs of weather conditions than any statistical model can hope to obtain. On the other hand, in a strong labor market, employers and employees may make greater efforts to overcome weather disruptions, leading to a problem of endogeneity with this measure.[8]

Table 2 reports the parameter estimates from specifications 1 through 8. Coefficients on snowfall are generally significantly negative, while coefficients on temperature are generally significantly positive, but only in the winter and early spring months. That is, unsurprisingly, unusually warm weather boosts employment (in these months), while unusually snowy weather lowers employment. The estimated coefficients give a "rule of thumb" for the effect of weather in month *t* on employment in month *t.* For example, in specification 1 we estimate that a 1-degree-Celsius decrease in average temperature in March lowers employment by 23,000.

Table 2 also reports the maximized log-likelihood from each specification, and *p* values from various likelihood ratio tests. We overwhelmingly reject a model with no weather effect in favor of specification 1. Among specifications 1 and 2 (using temperature or HDDs), the former gives the higher log-likelihood, so we prefer using temperature to HDDs. We reject specification 1 in favor of specifications 3 and 4, meaning that a snow indicator is important over and above the temperature effect. Among specifications 3 and 4, specification 4 (measuring snowfall using the RSI index) gives the higher log-likelihood, and this RSI index is consequently our preferred snowfall measure. The fact that the RSI index gives a better fit to employment than is obtained using simple snowfall totals indicates that Kocin and Uccellini (2004) and Squires and others (2014) succeeded in their aim of constructing indexes to measure the societal impact of snowstorms. However, we reject specification 4 in favor of specifications 5, 6, and 7, meaning that work absences, precipitation, and further lags are all important. Finally, there is no significant difference between specifications 4 and 8, meaning that there is not much evidence for weather on the 13th and 14th of the month having any additional impact.

7. Scientists agree that economic activity influences the climate, but this does not mean that it influences deviations of weather from seasonal norms.

8. Note also that there is a timing issue in using the CPS weather-related absences from work measure. That measure specifically refers to absence from work in the Sunday–Saturday period bracketing the 12th of the month. This lines up with the employment definition in the CES only for establishments with a Sunday–Saturday weekly pay period.

Table 2. Estimated Effects of Unusual Weather on Aggregate Employment

	Specification[a]							
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
γ_1	16.4**	−18.2**	12.6	13.8**	12.5*	13.7**	23.4***	12.3*
γ_2	33.6***	−38.6***	28.8***	23.3**	19.0**	22.6**	25.4***	23.4***
γ_3	23.3***	−26.8***	16.0**	18.3**	20.0***	19.0***	27.3***	17.9***
γ_4	−8.5	2.9	−18.1*	−6.3	−15.6*	−10.6	11.8	−10.3
γ_5	8.7	−4.1	20.7	12.3	16.8	16.3	28.6**	17.0
γ_6	22.7	55.0	24.4	22.3	24.6	15.9	6.4	15.0
γ_7	29.5	1,072	26.5	30.6	56.0	38.6	−6.4	28.9
γ_8	30.5	−183.4	26.3	30.3	44.5	29.5	18.1**	26.0
γ_9	6.5	−42.7	1.1	6.3	26.5	−11.2	12.5	12.0
γ_{10}	18.6*	−25.9*	14.0	16.7	23.6**	13.5	18.9*	20.3**
γ_{11}	25.2*	−36.3*	20.7	21.5	17.0	15.4	23.9	22.6**
γ_{12}	16.0*	−16.4	11.0	14.7	11.5	15.4	22.4**	13.0
γ_{13}			−7.62***	−37.74**	−20.36	−39.1**	−77.63***	−24.73*
γ_{14}					−0.29***	12.3**		
log-likelihood[b]	−1968.9	−1970.1	−1965.5	−1964.7	−1952.3	−1961.9	−1957.9	−1964.2

(*continued on next page*)

Table 2. Estimated Effects of Unusual Weather on Aggregate Employment (*Continued*)

Likelihood ratio tests[c]	*p values*	*Conclusion*
H_0 : No weather vs. Specification 1	0.00	Reject exclusion of temperature
H_0 : Specification 1 vs. Specification 3	0.01	Reject exclusion of snow
H_0 : Specification 1 vs. Specification 4	0.00	Reject exclusion of snow (RSI)
H_0 : Specification 4 vs. Specification 5	0.00	Reject exclusion of absences
H_0 : Specification 4 vs. Specification 6	0.02	Reject exclusion of precipitation
H_0 : Specification 4 vs. Specification 7	0.00	Reject exclusion of lags
H_0 : Specification 4 vs. Specification 8	0.59	Do not reject exclusion of 13th and 14th

Source: Authors' analysis, based on CES survey data.

a. The top panel of the table lists the parameter estimates from fitting specifications 1 through 8 (see text) to aggregate employment data. In all cases, $\gamma_1, \ldots, \gamma_{12}$ refer to the coefficients on the unusual temperature variable interacted with dummies for January to December, respectively (except heating degree days for specification 2). Meanwhile, γ_{13} refers to various snow effects (defined in text) and γ_{14} refers to the effects of seasonally adjusted self-reported work absences due to weather and precipitation in specifications 5 and 6, respectively. Statistical significance indicated at the *10 percent, **5 percent, and ***1 percent levels. Data units are as follows: Employment is measured in thousands, temperature is measured in degrees Celsius, snowfall (non-RSI) is measured in millimeters, snowfall (RSI) is measured in the scale that defines the index, precipitation is measured in millimeters, and work absences are measured in thousands.

b. This row gives the log-likelihood of each model. The specification with no weather effects at all has a log-likelihood of −1993.7.

c. Bottom panel of the table reports *p* values from various likelihood ratio tests comparing alternative specifications.

We considered some other specifications as well. First, we added the value of damage done by large hurricanes in the previous month,[9] relative to the 30-year average, to specification 4. However, this did not significantly improve specification 4, and so we do not consider hurricanes further.[10] Second, we amended specification 4 to allow for a nonlinearity, whereby positive and negative values of unexpected weather can have asymmetric effects. Again this did not significantly improve specification 4. Third, we modified specification 4 to use a weighted average of temperature in the nine different climate regions of the United States (as defined by the National Centers for Environmental Information), estimating the weights along with all the other parameters to maximize the likelihood of the national employment data. But this gave a barely significant improvement in likelihood, and the estimated weights were imprecisely estimated, and in some cases they were quite implausible in magnitude (notably, the northeast region received no weight at all). Clearly, weather conditions can differ greatly by region, but it does not seem that the separate effects of regional weather variation on national employment data are econometrically well identified.[11]

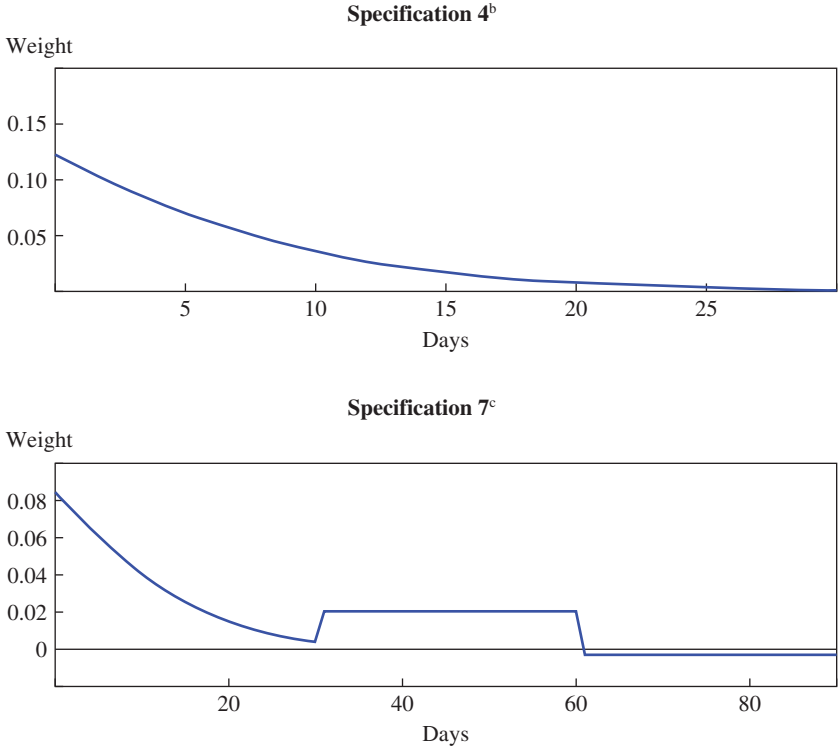
The upper panel of figure 1 plots the MIDAS polynomial implied by the pseudo-maximum likelihood estimates of a and b in specification 4. The estimated polynomial puts most weight on the few days up to and including the 12th of the month. This pattern can be found in the other specifications as well. The lower panel of figure 1 plots the lag structure $\{w_j\}_{j=0}^{90}$ corresponding to the estimates of specification 7. This specification allows for richer dynamics of the weather effect. The estimated value of c is positive, meaning that the weather effect in the level of employment lasts into the subsequent month. The estimated value of d is of very small magnitude but is negative. This means that the point estimates suggest that bad weather

9. This is the value in 2010 dollars, deflated by the price deflator for construction, as discussed in Blake, Landsea, and Gibney (2011).

10. We estimate that every billion dollars (in 2010 dollars) in unusual hurricane damage increases employment in that month by 287 jobs, with a 95 percent confidence interval of [–919, 1,493].

11. If one were instead trying to model regional employment data, then it would make sense to use regional weather data. However, as discussed earlier, the national employment data receive almost all of the focus in the media and among economists, policymakers and traders in financial markets, and these data cannot be built up from state level data. In addition, there may be spillover effects of weather in one region on economic activity in other regions, such as a large local snowstorm disrupting transportation between regions. Our equations fit national employment to national weather series in a parsimonious manner to allow for these potential effects.

Figure 1. Estimated MIDAS Polynomial[a]



Source: Authors' analysis.

a. Plots the weights w_j against j (in days) where parameters are set equal to their maximum likelihood estimates, fitting equation 1 to aggregate NSA employment in specifications 4 and 7. The weight for $j = 0$ is the weight attributed to unusual weather on the 12th day of the month (corresponding to the CES survey date).

b. In this panel, the underlying estimates of a and b are −3 and −2.01, respectively.

c. In this panel, the underlying estimates of a, b, c, and d are −1.77, −1.30, 0.02, and −0.003, respectively.

actually boosts employment two months later. This could be because of a catch-up effect. For example, if bad weather delayed a construction project in February, then this might make the builder employ more workers than otherwise in April to try to get back on schedule. A useful way of thinking of the lag structure in specification 7 is that if the average weight given to weather in the 30 days up to and including the 12th of the month[12] is

12. The weight given to the 30 days up to and including the 12th of the month is not constant—this is the average weight given to days in this window. The actual weights are shown in the lower panel of figure 1.

normalized to 1, then the weights given to weather in the previous two months are 0.6 and –0.1, respectively.

II. Weather and Seasonal Adjustment

The X-13 ARIMA[13] seasonal adjustment methodology, used by the BLS and other U.S. statistical agencies, is quite involved. Let y_t be a monthly series (possibly transformed) that is to be seasonally adjusted. The methodology first involves fitting a seasonal ARIMA model

$$(2) \qquad \phi(L)\Phi(L^{12})(1-L)^{d}(1-L^{12})^{D}(y_t - \beta' x_t) = \theta(L)\Theta(L^{12})\varepsilon_t,$$

where x_t is a vector of user-chosen regressors, β is a vector of parameters, L denotes the lag operator, $\phi(L)$, $\Phi(L^{12})$, $\theta(L)$, and $\Theta(L^{12})$ are polynomials of orders p, P, q, and Q, respectively, d and D are integer difference operators, and ε_t is an independent and identically distributed error term. The model is estimated by maximum likelihood. The regression residuals, $y_t - \hat{\beta}' x_t$, are then passed through filters, as described in the appendix of Jonathan Wright (2013) and in more detail in Dominique Ladiray and Benoît Quenneville (2001), to estimate seasonal factors. Note that our specifications in the previous section are all special cases of equation 2.

Seasonal adjustment in the CES is implemented at the three-digit NAICS[14] level (or more disaggregated for some series), and these series are then aggregated to construct seasonally adjusted total nonfarm payrolls. In all, there are 150 disaggregates. We used the modeling choices, including ARIMA lag orders in equation 2, chosen by the BLS for each of the disaggregates, but simply included measures of unusual weather, x_t^w, in the vector of user-chosen regressors, x_t. We consider the specifications in the previous section. Depending on the specification, our weather regressor x_t^w consists of the unusual temperature for month t, as constructed in the previous section,[15] interacted with 12 monthly dummies, the unusual snowfall for month t (defined analogously, but not interacted with any dummies), and/or abs_t. All in all, this gives a total of 12 to 14 elements

13. ARIMA stands for autoregressive integrated moving average.

14. North American Industry Classification System.

15. In specification 1 for aggregate employment data, let $\hat{a}$ and $\hat{b}$ denote pseudo-maximum likelihood estimates of a and b. We measure the unusual temperature for month t as $\sum_{j=0}^{30} B\left(\frac{j}{30}, \hat{a}, \hat{b}\right) temp^*_{s-j}$, where $temp^*_s$ is the unusual temperature on the 12th day of month t.

in x_t^w, depending on the specification, for inclusion as regressors in the X-13 filter. As in the previous section, we are assuming that the effect of weather is linear.

The sample period is January 1990 to May 2015 in all cases—the sample period is dictated by the fact that January 1990 is the start date for many of the 150 employment disaggregates.[16] For each of the 150 series, we compute the seasonally adjusted data net of weather effects, which we refer to as seasonally-and-weather-adjusted (SWA). It is important to note that when we construct the SWA data we remove the weather effects before computing the seasonal adjustment and we do not add back these effects. In contrast, when the BLS judgmentally adjusts for extreme weather effects before calculating seasonal adjustments, it adds back these initial adjustments. The BLS's aim is not to purge the data of weather effects, but simply to ensure that the unusual weather does not contaminate estimates of seasonal patterns. Our aim for making weather adjustments is not only to improve seasonal adjustment but also to produce data that are purged of unusual weather effects. A researcher could follow our methodology and then add the weather effects back in, which would keep the weather effects in the data but not let them affect seasonal patterns.[17] But in this paper, we control for both the direct effect of weather on the data and the impact of weather on seasonal adjustment. The resulting SWA data can then be summed across the 150 disaggregates and can be compared with the standard version of data that are only seasonally adjusted (SA).[18]

The idea of preventing unusual weather from affecting seasonal factors is a little tricky in the presence of climate change, because unusual weather might change one's beliefs about seasonal norms. However, climatologists measure seasonal norms from 30-year averages (World Meteorological Organization 2011), whereas the X-13 filter effectively estimates seasonal factors from averaging just a few years' data. Allowing unusual

16. Our weather data go back to 1960, allowing us to measure unusual weather by subtracting off a backward-looking 30-year average.

17. This is not what the BLS currently does. The BLS adjusts for specific extreme weather events before computing seasonal factors on a case-by-case basis, rather than doing so automatically as we envision.

18. Our SA data differ somewhat from the official SA data because we use current-vintage data and the current specification files. In contrast, the official seasonal factors in the CES are frozen as estimated five years after the data are first released. Also, we use the full sample back to 1990 for seasonal adjustment. Nevertheless, our SA and SWA data are completely comparable.

weather to affect seasonal factors as estimated in the X-13 makes them too volatile.[19]

Note also that our methodology uses aggregate employment to estimate the parameters *a, b, c,* and *d* that specify how employment is affected by the weather on different days. However, the seasonal-and-weather adjustment is otherwise conducted by applying the full X-13 methodology at the disaggregate level, as described earlier. Other than these parameters (which affect the construction of the monthly weather regressors x_t^w), no parameters from the estimation of equation 1 are used in our seasonal-and-weather adjustment. We use the same lag weights and model specification for each of the disaggregates for reasons of computational cost, parsimony, and ease of interpretation. The price that we pay for this is that we do not allow the persistence of weather effects or the choice of weather indicators to differ across industries. It is important to emphasize that we do allow the magnitude of weather effects to differ across industries—we only restrict the lag structure and choice of weather indicators to be the same.

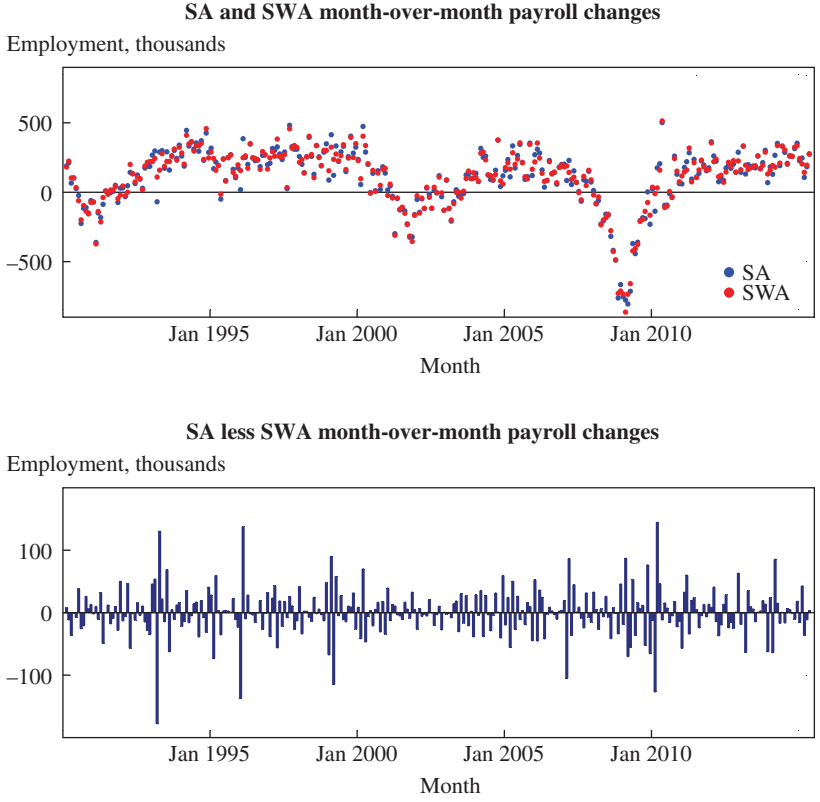
II.A. Results of Specification 4

We start by considering specification 4 as the baseline case for constructing the weather variables that are used in equation 2 for 150 CES disaggregates. We believe that temperature and snowfall capture a large fraction of the potential weather effects, and specification 4 includes both temperature and snowfall effects in a straightforward manner, with snowfall measured using the RSI index. Results from using other specifications are discussed in subsection II.B.

Figure 2 compares total nonfarm payrolls using ordinary seasonal adjustment and our seasonal-and-weather adjustment, using this specification. The top panel shows the month-over-month changes in total payrolls with ordinary seasonal adjustment along with the comparable series that we constructed by adjusting for both abnormal weather and normal seasonal patterns. The bottom panel shows the differences in the two series (ordinary SA less SWA). The differences represent the combination of the directly estimated weather effects that are removed from the SWA series and differences between the seasonal factors in the two series. The latter source of

19. Even preventing unusual weather from affecting seasonal factors, the seasonal factors will eventually catch up to climate change because we define unusual weather relative to a rolling 30-year average.

Figure 2. Difference between SA and SWA Month-over-Month Payroll Changes[a]



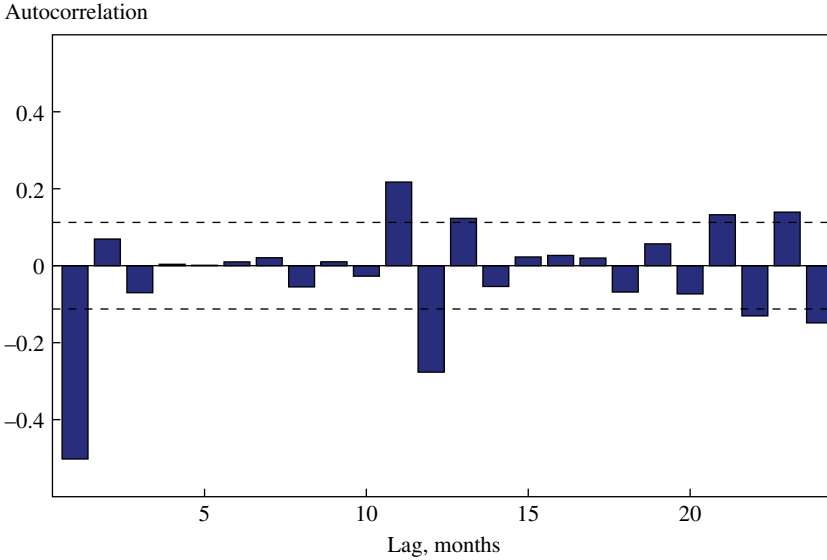
Source: Authors' analysis, based on CES survey data.

a. Shows the month-over-month change in total nonfarm payrolls using ordinary seasonal adjustment less the corresponding change using seasonal-and-weather adjustment. This shows the estimated effect of the weather, including the effect of controlling for the weather on seasonal factors. The exercise uses temperature interacted with month dummies and RSI snowfall as weather variables (corresponding to specification 4).

differences is driven by the fact that failing to control for unusual weather events affects estimated seasonal factors.

Of course, the weather effects in the bottom panel of figure 2 can be either positive or negative. They can be more than 100,000 in absolute magnitude. While these effects are generally small relative to the sampling error in preliminary month-over-month payroll changes in the CES (standard deviation of 57,000), financial markets, the press, and the Federal Reserve are hypersensitive to employment data. The weather adjustments that we propose might often substantially alter their perceptions of the labor market.

Figure 3. Autocorrelation of Weather Effects[a]



Source: Authors' analysis, based on CES survey data.

a. Shows the sample autocorrelation function of weather effects, defined as the month-over-month change in total nonfarm payrolls using ordinary seasonal adjustment less the corresponding change using seasonal-and-weather adjustment. The horizontal dashed lines are the critical values for sample autocorrelations to be statistically significant at the 5 percent level. See note to figure 2.

AUTOCORRELATION Figure 3 shows the autocorrelogram of estimated weather effects. At a lag of one month, the weather effects are significantly negatively autocorrelated. This is because they are estimates of the weather effects in month-over-month changes. Unusually cold weather in month t will lower the change in payrolls during that month, but will boost the change in payrolls for month $t + 1$, assuming that normal weather returns in month $t + 1$.

The autocorrelation of the weather effect in payroll changes at lag 12 is also significantly negative. This is because bad weather has some effect on estimated seasonal factors, leading to an "echo" effect of the opposite sign one year later.[20] This underscores the importance of integrating the

20. Wright (2013) argues that the job losses in the winter of 2008–09 produced an echo effect of this sort in subsequent years. The distortionary effects of the Great Recession on seasonals are of course far bigger than the effects of any weather-related disturbances.

weather adjustment into the seasonal adjustment process, as opposed to simply attempting to control for the effect of weather on data that have been seasonally adjusted in the usual way.

RECENT WINTERS In figure 2, the effects of the unusually cold winter of 2013–14 can be seen. We estimate that weather effects lower the month-over-month payroll change for December 2013 by 62,000 and by 64,000 in February 2014. Meanwhile, we estimate that the weather effect raised the payroll change for March 2014 by 85,000 as more normal weather returned. The weather effect was quite consequential, but still does not explain all of the weakness in employment reports during the winter of 2013–14. In March 2015, colder-than-normal weather is estimated to have lowered monthly payroll changes by 36,000.

HISTORICAL EFFECTS The winters of 2013–14 and 2014–15 are far from the biggest weather effects in the sample. The data in February and March 2007 contained a large swing, because that February was colder than usual. That fact was not missed by the Federal Reserve's Greenbook, which noted in March 2007 that

> in February, private nonfarm payroll employment increased only 58,000, as severe winter weather likely contributed to a 62,000 decline in construction employment.[21]

Payroll changes were weak in April and May 2012. Then–Federal Reserve chairman Ben Bernanke (2012), in testimony to the Joint Economic Committee, attributed part of this to weather effects, noting that

> the unusually warm weather this past winter may have brought forward some hiring in sectors such as construction where activity normally is subdued during the coldest months; thus, some of the slower pace of job gains this spring may have represented a payback for that earlier hiring.

The data in February and March 1999 also contained a big swing, since that February was unseasonably mild. According to our estimates, weather drove the month-over-month change in payrolls up by 90,000 in February 1999 and down by 115,000 the next month. The biggest effect in the sample was March 1993, when weather is estimated to have lowered employment growth by 178,000.[22] This is an enormous estimated weather effect, but it does not seem unreasonable: In March 1993, reported nonfarm

21. See page II-1 of the Federal Reserve's 2007 Greenbook here: http://www.federalreserve.gov/monetarypolicy/files/FOMC20070321gbpt220070314.pdf.

22. Note that there were very big snowstorms in three regions of the country in that month.

Table 3. Weather Effect in Monthly Payroll Changes, Top 10 Absolute Effects[a]

Month	*Weather effect*
March 1993	−178
March 2010	+144
February 1996	+137
January 1996	−137
April 1993	+130
February 2010	−127
March 1999	−115
February 2007	−105
February 1999	+90
March 2007	+87

Source: Authors' analysis, based on CES survey data.

a. Shows the difference in monthly payroll changes (in thousands) that are SA less those that are SWA, for the 10 months where the effects are biggest in absolute magnitude. These are constructed by applying either the seasonal adjustment or the seasonal-and-weather adjustment to all 150 CES disaggregates, and then adding them up, as described in the text. The exercise uses temperature interacted with month dummies and RSI snowfall as weather variables (corresponding to specification 4).

payrolls fell by 49,000, while employment growth was robust in the previous and subsequent few months.[23]

Table 3 lists the 10 months in which the weather effect (the bottom panel of figure 2) is the largest in absolute magnitude. These all occur in the first four months of the year. They turn out to be five pairs of adjacent months as the effects of unusual weather are followed by bounce-backs when more seasonal weather returns.

Table 4 gives the minimum, maximum, and standard deviation of the total weather effect in payroll changes broken out by month.[24] The standard deviation is the largest in March (68,000), followed by February (58,000). The standard deviations show that weather effects are potentially economically significant in winter and early spring but are relatively small in the summer months.

Figure 4 plots the difference between ordinary SA data and SWA data for payroll changes in the construction sector alone (again using specification 4). Weather effects in the construction sector drive a bit less than half of total weather effects.

23. These are current-data-vintage numbers, with ordinary seasonal adjustment. The first released number for March 1993 was −22,000. The BLS employment situation write-up for that month made reference to the effects of the weather. But the BLS made no attempt to quantify the weather effect.

24. Means are not shown because they are close to zero by construction.

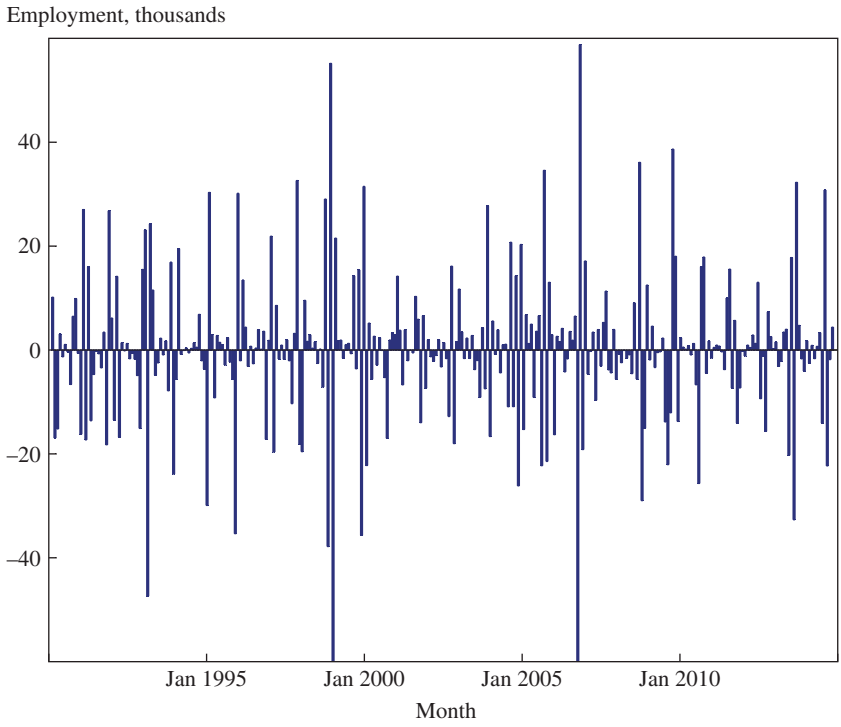
Table 4. Weather Effect in Monthly Payroll Changes, Summary Statistics[a]

Month	*Standard deviation*	*Minimum*	*Maximum*
January	42	–137	53
February	58	–127	137
March	68	–178	144
April	44	–57	130
May	24	–49	53
June	17	–36	27
July	22	29	69
August	18	–63	17
September	15	–24	31
October	20	–52	32
November	26	–40	76
December	38	–66	63
Overall	36	–178	144

Source: Authors' analysis, based on CES survey data.

a. Shows the standard deviation, minimum, and maximum of the monthly payroll changes (in thousands) that are SA less those that are SWA adjusted, broken out by month. See note to table 3.

Figure 4. Difference between SA and SWA Month-over-Month Payroll Changes in Construction Sector[a]



Source: Authors' analysis, based on CES survey data.

a. See note to figure 2.

In all, the weather adjustment involves estimating 14 parameters in β^w for each of the 150 disaggregates for a total of 2,100 parameters. We do not report all of these parameter estimates. Most of the parameters are individually statistically insignificant, but the parameters associated with temperature in December, January, February, and March, as well as the parameters associated with snowfall, are significantly negative for components of construction employment.

We deliberately decided against a strategy of setting parameter estimates that are individually insignificant to zero. In general, assuming that a parameter is precisely zero because it is not statistically significant seems a dubious approach, and this may be particularly true when doing a bottom-up adjustment for weather effects. For an individual disaggregate, a weather effect might be minor, but these weather effects are likely to be positively correlated across disaggregates, and so the weather effect might be much more important in the aggregate data that we ultimately care about.

PERSISTENCE Purging employment data of the weather effect might make the resulting series more persistent, in much the same way as purging consumer price index inflation of the volatile food and energy component makes the resulting core inflation series smoother, as discussed in the introduction. To investigate this, we compare the standard deviation and autocorrelation of month-over-month changes in SA and SWA payroll data, both for total payrolls and for nine industry subaggregates. The results are shown in table 5.

In the aggregate, month-over-month payroll changes show a higher degree of autocorrelation using SWA data than using SA data. This primarily reflects the fact that the weather adjustments remove noise from the levels data which is a source of negative autocorrelation in month-over-month changes. In fact, in every sector except government, payroll changes show a higher degree of autocorrelation using SWA data than using SA data. The effect is small in most sectors, with the exception of construction, where the proposed weather adjustment raises autocorrelation from 0.59 to 0.77. Particularly in the construction sector, weather adjustment removes noise that is unrelated to the trend, cyclical, or seasonal components. This gives a better measure of the underlying strength of the economy.

II.B. Results with Other Specifications

We also considered the effects on seasonal adjustments from using other specifications discussed in section I. In particular, we considered specifications 5, 6, 7, and 8 as alternatives to specification 4. Specification 5 includes absences from work, specification 6 includes precipitation, specification 7

Table 5. Autocorrelation and Standard Deviation of Month-over-Month Changes in SA and SWA Nonfarm Payroll Data, by Sector[a]

	Autocorrelation		*Standard deviation*	
Sector	*SA data*	*SWA data*	*SA data*	*SWA data*
Mining and logging	0.662	0.686	5.1	5.0
Construction	0.586	0.768	39.0	35.9
Manufacturing	0.739	0.756	50.4	50.2
Trade, transportation, and utilities	0.631	0.651	53.2	52.7
Information	0.625	0.645	23.2	23.0
Professional and business services	0.572	0.609	53.7	52.9
Leisure and hospitality	0.324	0.374	28.6	27.2
Other services	0.496	0.533	8.9	8.8
Government	0.036	0.034	51.5	51.2
Total	0.800	0.840	214.4	210.7

Source: Authors' analysis, based on CES survey data.

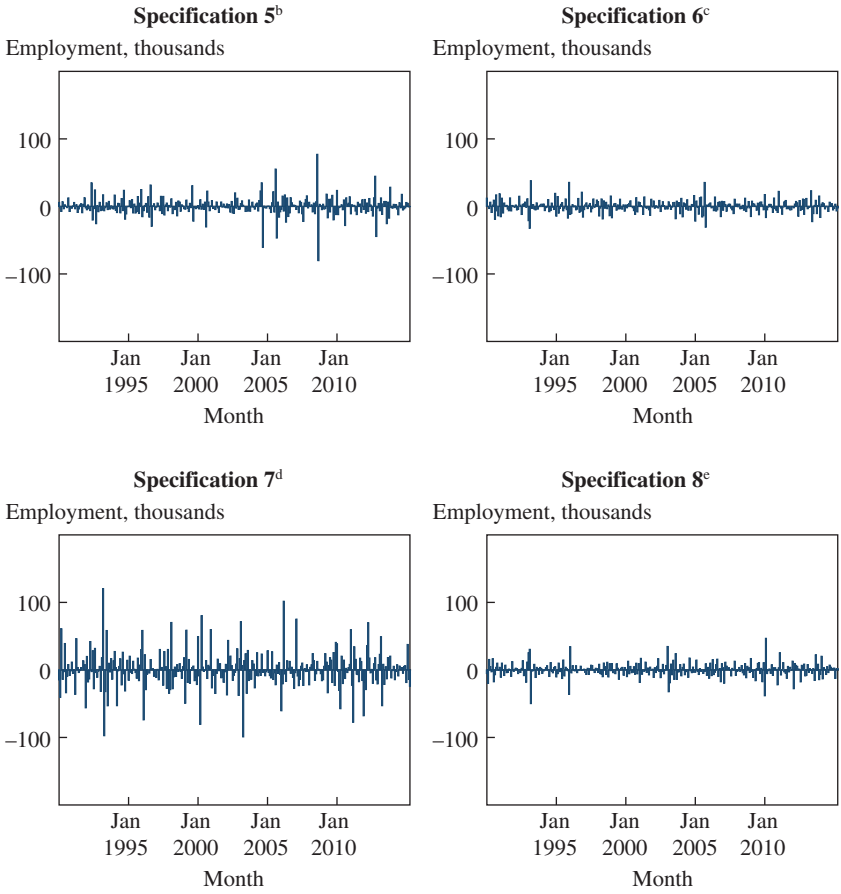
a. Reports the first-order autocorrelation and standard deviation of seasonally adjusted (SA) month-over-month payroll changes (in thousands; total and by industry) and of the corresponding seasonally-and-weather-adjusted (SWA) data. The exercise uses temperature interacted with month dummies and RSI snowfall as weather variables (corresponding to specification 4).

adds monthly lags to admit richer dynamics, and specification 8 includes weather on the 13th and 14th of the month. Figure 5 shows the difference between SWA data in each of these specifications and the SWA data in specification 4 (that simply used temperature and the RSI index). These charts show that only specification 7 produces noticeably different results. Since the more complicated models make little difference to the weather adjustment, and since simpler models are easier to understand, we prefer specification 4 to specifications 5, 6, and 8.[25]

Including monthly lags (specification 7) does, however, make a material difference to SWA data, and so we do think of this as an alternative benchmark approach to weather adjustment. Specification 4 forces the effects of unusual weather on the level of employment to disappear the next month, whereas specification 7 is more flexible regarding the dynamics of weather effects. Figure 6 shows the difference between month-over-month payroll changes using ordinary seasonal adjustment and SWA data using specification 7. The weather effects for changes in employment are still negatively autocorrelated, but they are much less so when using lags; the first autocorrelation is –0.5 in specification 4, but –0.2 in specification 7.

25. While including absences from work in specification 5 seldom makes a material difference, an exception is September 2008. In this month, the number who reported absence from work due to weather spiked to levels normally observed only in winter. We speculate that this might owe to the fact that Hurricane Ike was moving toward Texas during the survey week.

Figure 5. Difference between SWA Month-over-Month Payroll Changes in Alternative Specifications[a]



Source: Authors' analysis, based on CES survey data.

a. The four subpanels of this figure show the month-over-month payroll changes using SWA data, where the weather variables are as in specifications 5, 6, 7, and 8, respectively, less the corresponding SWA data using specification 4. The figure shows the incremental effects of each of these additions to the specification on SWA data.

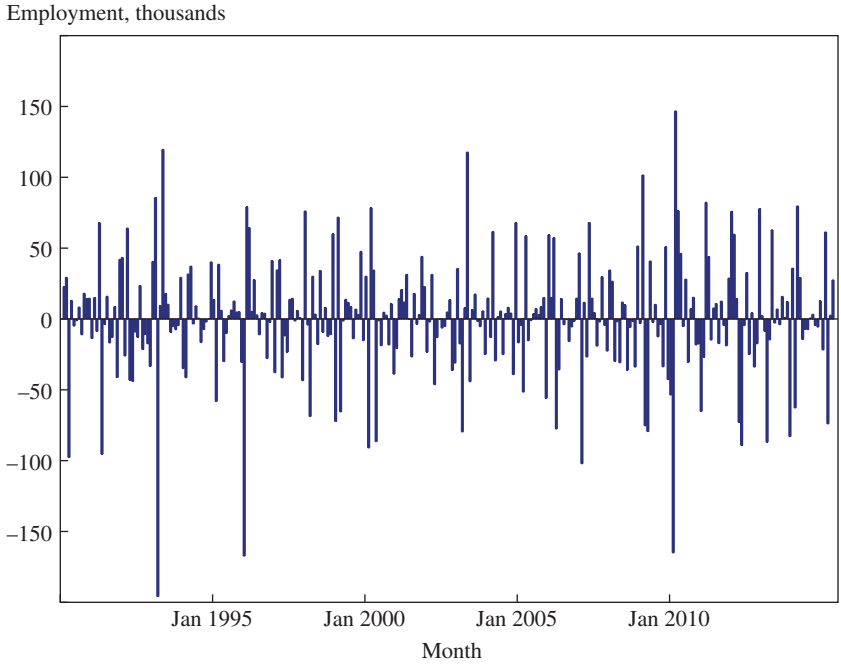
b. Relative to specification 4, specification 5 adds CPS work absences due to weather.

c. Relative to specification 4, specification 6 instead adds precipitation.

d. Relative to specification 4, specification 7 instead adds two monthly lags.

e. Relative to specification 4, specification 8 adds weather on the 13th and 14th of the current month.

Figure 6. Difference between SA and SWA Month-over-Month Payroll Changes Using Specification 7[a]



Source: Authors' analysis, based on CES survey data.
a. See note to figure 2. In this figure, lags of weather indicators in the previous two months are also included (as in specification 7).

Table 6 lists the 10 months in which the weather effects from using this specification are largest in absolute magnitude. Only 5 of these months are also found in table 3. It is interesting to note that table 6 includes only one pair of adjacent months (February and March 2010), while all of the months in table 3 are paired with an adjacent month, which is not entirely surprising because the bounce-back phenomenon from specification 7 is weaker. We computed analogs of tables 4 and 5 for specification 7, but they are similar to the original tables so we do not include them in the paper.

III. NIPA Data

Our focus in this paper has been on the employment report, both because it is the most widely followed economic news release and because it is possible to closely replicate the seasonal adjustment process that the BLS uses

Table 6. Weather Effect on Monthly Payroll Changes, Top 10 Absolute Effects Using Specification 7[a]

Month	*Weather effect*
March 1993	−196
January 1996	−167
February 2010	−165
March 2010	+147
May 1993	+120
May 2003	+118
February 2007	−102
February 2009	+102
April 1990	–98
May 1991	–95

Source: Authors' analysis, based on CES survey data.

a. Shows the monthly payroll changes (in thousands) that are SA less those that are SWA, for the 10 months where the effects are biggest in absolute magnitude. These are constructed by applying either the seasonal adjustment or the seasonal-and-weather adjustment to all 150 CES disaggregates, and then adding them up, as described in the text. The exercise uses temperature interacted with month dummies and RSI snowfall along with two monthly lags as weather variables (corresponding to specification 7).

in the reported CES data. GDP and other NIPA-based economic data are also widely followed and are potentially subject to weather effects. In fact, weather effects could be more important for these series, because harsh weather only affects employment statistics when it causes an employee to miss an entire pay period, but it could have broader effects on NIPA series by lowering hours worked or consumer spending. On the other hand, weather effects on NIPA series could be mitigated by the fact that NIPA data are averaged over a whole quarter, not just a pay period.

III.A. NIPA Weather Adjustment

Unfortunately, the SWA steps described in the previous section cannot be applied to NIPA data because there is no way for researchers to replicate the seasonal adjustment process in these data, let alone to add weather effects to it.[26]

26. Although the BEA compiles NIPA data, seasonal adjustment is done at a highly disaggregated level, and many series are passed from other agencies to the BEA in seasonally adjusted form. As noted in Wright (2013) and Manski (2015), while the BEA used to compile not seasonally adjusted NIPA data, they stopped doing so a few years back as a cost-cutting measure. Happily, the June 2015 Survey of Current Business indicated plans to resume publication of not seasonally adjusted aggregate data, but this will still not allow researchers to replicate the seasonal adjustment process.

Table 7. Coefficient Estimates for Equation 3, 1990Q1–2015Q2[a]

	Real GDP	*Personal consumption*	*Private investment*	*Government expenditures*	*Exports*	*Imports*
γ_1	0.08***	0.04**	0.19	0.06*	0.26**	0.15*
	(0.03)	(0.02)	(0.12)	(0.03)	(0.11)	(0.09)
γ_2	0.11**	0.06	0.29	–0.08	0.28	0.09
	(0.05)	(0.05)	(0.28)	(0.06)	(0.18)	(0.13)
γ_3	0.04	0.01	–0.33	0.07	0.08	–0.27
	(0.04)	(0.05)	(0.37)	(0.05)	(0.23)	(0.19)
γ_4	0.05	0.02	–0.09	0.07	0.12	–0.10
	(0.04)	(0.04)	(0.22)	(0.05)	(0.14)	(0.11)
γ_5	0.22	–0.04	7.28*	–2.83**	0.68	–1.21
	(0.80)	(0.57)	(4.17)	(1.41)	(2.90)	(2.85)

Source: Authors' analysis, based on September 2015 vintage NIPA data.

a. Data units are as follows: NIPA growth rates are measured in annualized percentage points, temperature is measured in degrees Celsius, and snowfall is measured in millimeters. Standard errors in parentheses. Statistical significance indicated at the *10 percent, **5 percent, and ***1 percent levels.

As an alternative, we instead apply weather adjustments directly to seasonally adjusted NIPA aggregates. We consider the model

$$(3)\quad y_t = \mu_1 s_{1t} + \mu_2 s_{2t} + \mu_3 s_{3t} + \mu_4 s_{4t} + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4}$$
$$+ \gamma_1 w_{1t} d_{1t} + \gamma_2 w_{1t} d_{2t} + \gamma_3 w_{1t} d_{3t} + \gamma_4 w_{1t} d_{4t} + \gamma_5 (w_{2t} - w_{2t-1}) + \varepsilon_t,$$

where y_t is the quarter-over-quarter growth rate of real GDP or some component thereof, $s_{1t}, \ldots, s_{4t}$ are four quarterly dummies,[27] w_{1t} is the unusual temperature in quarter t (defined as the simple average of daily values in that quarter), w_{2t} is the unusual snowfall in quarter t (using the RSI index), and $d_{1t}, \ldots, d_{4t}$ are four quarterly variables, each of which takes on the value 1 in a particular quarter, –1 in the next quarter, and 0 otherwise. The particular specification in equation 3 has the property that no weather shock can ever have a permanent effect on the level of real GDP—any weather effect on growth has to be "paid back" eventually, although not necessarily in the subsequent quarter, given the lagged dependent variables.[28] Our sample period is 1990Q1–2015Q2, using September 2015 vintage data. Coefficient estimates are shown in table 7 for real GDP growth and selected

27. The inclusion of these quarterly dummies is motivated by "residual seasonality" discussed further below.

28. Macroeconomic Advisers (2014) find that snowfall effects on growth are followed by effects of opposite sign and roughly equal magnitude in the next quarter.

Table 8. Adjustments to NIPA Variable Growth Rates in 2015[a]

	Quarter	*SA data*[b]	*SWA data*[c]	*SSWA data*[d]
Real GDP	Q1	0.6	1.5	3.3
	Q2	3.9	3.1	2.6
Personal consumption	Q1	1.7	2.0	2.4
	Q2	3.6	3.2	3.4
Private investment	Q1	8.6	9.6	12.7
	Q2	5.0	3.1	1.0
Government expenditures	Q1	–0.1	0.6	0.9
	Q2	2.6	2.4	1.3
Exports	Q1	–6.0	–3.6	2.2
	Q2	5.1	3.0	1.0
Imports	Q1	7.1	8.4	8.4
	Q2	3.0	2.2	1.7

Source: Authors' analysis, based on September 2015 vintage NIPA data.

a. Shows the quarter-over-quarter growth rates of real GDP and its five components in 2015Q1 and 2015Q2. All entries are in annualized percentage points.

b. Refers to seasonally adjusted data published by the BLS.

c. Refers to seasonally-and-weather-adjusted data using the method described in section III.

d. Refers to seasonally-and-weather-adjusted data, as described in section III, with a second round of seasonal adjustment applied using the X-13 default settings.

components. For real GDP growth, unusual temperature is statistically significant in the first and second quarters.

We think that the assumption that no weather shock can have a permanent effect on the level of GDP is an important and reasonable restriction to impose. Nevertheless, we tested this restriction. We ran a regression of y_t on four quarterly dummies, four lags of y_t, unusual temperature interacted with quarterly dummies, lags of unusual temperature interacted with quarterly dummies, unusual snowfall, and lagged unusual snowfall. In this specification, there were 18 free parameters—equation 3 is a special case of this, imposing five constraints that can be tested by a likelihood ratio test. The restriction is not rejected at the 5-percent level for GDP growth or any of the components, except government spending where the p value is 0.04.

Having estimated equation 3, we then compute the dynamic weather effect by comparing the original series to a counterfactual series where all unusual weather indicators are equal to zero ($w_{1t} = w_{2t} = 0$), but with the same residuals. The difference between the original and counterfactual series is our estimate of the weather effect.

Table 8 shows the quarter-over-quarter growth rates of real GDP and components in 2015Q1 and 2015Q2 both in the data as reported and after our proposed weather adjustment. Weather adjustment raises the estimate of growth in the first quarter from 0.6 percentage point at an annualized

rate to 1.4 percentage points. However, the estimate of growth in the second quarter is lowered from 3.7 to 2.8 percentage points. Weather adjustment makes the acceleration from the first quarter to the second quarter less marked.

III.B. Residual Seasonality

Our paper is about the effects of weather on economic data, not seasonal adjustment. But an unusual pattern has prevailed for some time in which first-quarter real GDP growth is generally lower than growth later in the year, raising the possibility of "residual seasonality"—the Bureau of Economic Analysis (BEA)'s reported data may not adequately correct for regular calendar-based patterns. This is a factor, separate from weather, that might have lowered reported growth in 2015Q1. Glenn Rudebusch, Daniel Wilson, and Tim Mahedy (2015) apply the X-12 seasonal filter to reported seasonally adjusted aggregate real GDP and find that their "double adjustment" of GDP makes a substantial difference.[29]

The BEA has subsequently revisited its seasonal adjustment and made changes in the July 2015 annual revision. The changes might have mitigated residual seasonality, but it is important to note that the BEA has not published a complete historical revision to GDP and its components, instead only reporting improved seasonally adjusted data starting in 2012. We did an exercise in the spirit of Rudebusch, Wilson, and Mahedy (2015) by taking our weather-adjusted aggregate real GDP (and components) data and putting them through the X-13 filter. This double seasonal adjustment is admittedly an ad hoc procedure, especially given that BEA uses a different seasonal adjustment method for data after 2012 than for data before 2012; consequently, we treat our procedure's results with particular caution. Nonetheless, the resulting growth rates in the first two quarters of 2015 are also shown in table 7. After these two adjustments, growth was quite strong in the first quarter, but weaker in the second quarter, which is the opposite of the picture one obtains using published data. It is interesting to note that the "double seasonal adjustment" has an especially large effect on investment and exports, suggesting that these are two areas in which seasonal adjustment procedures might benefit from further investigation.

29. On the other hand, Gilbert and others (2015) find no statistically significant evidence of residual seasonality. The two papers are asking somewhat different questions. Gilbert and others (2015) are asking a testing question, and, while the hypothesis is not rejected, the *p* values are right on the borderline despite a short sample. Rudebusch, Wilson, and Mahedy (2015) are applying an estimation methodology.

IV. Conclusion

Seasonal effects in macroeconomic data are enormous. These seasonal effects reflect, among other things, the consequences of regular variation in weather over the year. However, the seasonal adjustments that are applied to economic data are not intended to address deviations of weather from seasonal norms. Yet these weather deviations have material effects on macroeconomic data. Recognizing this fact, this paper has operationalized an approach for simultaneously controlling for both normal seasonal patterns and unusual weather effects. Our main focus has been on monthly employment data in the CES, or the "establishment survey." The effects of unusual weather can be very important, especially in the construction sector and in the winter and early spring months. Monthly payroll changes are somewhat more persistent for seasonally-and-weather adjusted data than for ordinary seasonally adjusted data, suggesting that this gives a better measure of the underlying momentum of the economy.

The physical weather indicators considered in this paper are all available on an almost real-time basis—the reporting lag is inconsequential. The National Centers for Environmental Information make daily summaries for 1,600 stations available with a lag of less than 48 hours. In addition, the regional snowfall impact indexes that we use are typically computed and reported within a few days after a snowstorm ends. One weather indicator that we considered is the number of absences from work due to weather. This has a somewhat longer publication lag, but by construction is still available at the time of the employment report.

It would be good if weather adjustments of this sort could be implemented by statistical agencies as part of their regular data reporting process. Because they have access to the underlying source data, they have more flexibility in doing so than the general public—for example, some of the 150 disaggregates in the CES are not available until the first revision. Statistical agencies want data construction to use transparent methods that avoid ad hoc judgmental interventions, and that can be done for weather adjustment. U.S. statistical agencies nevertheless face severe resource constraints, and weather adjustment might well have an insufficiently high priority. In that case, weather adjustment could be implemented by end users of the data. We do not think weather-adjusted economic data should ever replace the underlying existing data, but as this paper demonstrates, weather adjustment can be a useful supplement to measure underlying economic momentum.

ACKNOWLEDGMENTS We are grateful to Katharine Abraham, Roc Armenter, Bob Barbera, Mary Bowler, François Gourio, Claudia Sahm, Tom Stark, and the editors for helpful discussions, and to Natsuki Arai for outstanding research assistance. All errors are our sole responsibility. The views expressed here are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

References

Andreou, Elena, Eric Ghysels, and Andros Kourtellos. 2010. "Regression Models with Mixed Sampling Frequencies." *Journal of Econometrics* 158, no. 2: 246–61.

Bernanke, Ben S. 2012. "Economic Outlook and Policy." Testimony before the Joint Economic Committee, U.S. Congress, June 7. http://www.federalreserve.gov/newsevents/testimony/bernanke20120607a.htm

Blake, Eric S., Christopher W. Landsea, and Ethan J. Gibney. 2011. "The Deadliest, Costliest, and Most Intense United States Tropical Cyclones from 1851 to 2010 (and Other Frequently Requested Hurricane Facts)." NOAA Technical Memorandum NWS NHC-6. Miami: National Weather Service, National Hurricane Center.

Bloesch, Justin, and François Gourio. 2014. "The Effect of Winter Weather on U.S. Economic Activity." *Economic Perspectives* 39, no. 1: 1–20.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. 2012. "Temperature Shocks and Economic Growth: Evidence from the Last Half Century." *American Economic Journal: Macroeconomics* 4, no. 3: 66–95.

Faust, Jon, and Jonathan H. Wright. 2013. "Forecasting Inflation." In *Handbook of Economic Forecasting* 2A, edited by Graham Elliott and Allan Timmermann. Amsterdam: North-Holland.

Fisher, R. A. 1925. "The Influence of Rainfall on the Yield of Wheat at Rothamsted." *Philosophical Transactions of the Royal Society of London (Series B)* 213: 89–142.

Foote, Christopher L. 2015. "Did Abnormal Weather Affect U.S. Employment Growth in Early 2015?" Current Policy Perspectives no. 15-2. Federal Reserve Bank of Boston.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov. 2004. "The MIDAS Touch: Mixed Data Sampling Regression Models." Working Paper. http://docentes.fe.unl.pt/~psc/MIDAS.pdf

———. 2005. "There Is a Risk-Return Trade-Off after All." *Journal of Financial Economics* 76, no. 3: 509–48.

Gilbert, Charles E., Norman J. Morin, Andrew D. Paciorek, and Claudia R. Sahm. 2015. "Residual Seasonality in GDP." FEDS Notes. Washington: Board of Governors of the Federal Reserve System.

Kocin, Paul J., and Louis W. Uccellini. 2004. "A Snowfall Impact Scale Derived from Northeast Snowfall Distributions." *Bulletin of the American Meteorological Society* 85, no. 2: 177–94.

Ladiray, Dominique, and Benoît Quenneville. 2001. *Seasonal Adjustment with the X-11 Method.* New York: Springer.

Macroeconomic Advisers. 2014. "Elevated Snowfall Reduced Q1 GDP Growth 1.4 Percentage Points." Blog post. http://www.macroadvisers.com/2014/04/elevated-snowfall-reduced-q1-gdp-growth-1-4-percentage-points/

Manski, Charles F. 2015. "Communicating Uncertainty in Official Economic Statistics: An Appraisal Fifty Years after Morgenstern." *Journal of Economic Literature* 53, no. 3: 1–23.

Miguel, Edward, Shanker Satyanath, and Ernest Serengeti. 2004. "Economic Shocks and Civil Conflict: An Instrumental Variables Approach." *Journal of Political Economy* 112, no. 4: 725–53.

Rudebusch, Glenn D., Daniel Wilson, and Tim Mahedy. 2015. "The Puzzle of Weak First-Quarter GDP Growth." Economic Letter no. 2015-16. Federal Reserve Bank of San Francisco.

Squires, Michael F., Jay H. Lawrimore, Richard R. Heim Jr., David A. Robinson, Mathieu R. Gerbush, and Thomas W. Estilow. 2014. "The Regional Snowfall Index." *Bulletin of the American Meteorological Society* 95, no. 12: 1835–48.

World Meteorological Organization. 2011. "Guide to Climatological Practices." WMO-No. 100. Geneva.

Wright, Jonathan H. 2013. "Unseasonal Seasonals?" *Brookings Papers on Economic Activity,* Fall: 65–110.

Yellen, Janet L. 2014. "The Economic Outlook." Testimony before the Joint Economic Committee, U.S. Congress, May 7. http://www.federalreserve.gov/newsevents/testimony/yellen20140507a.htm

———. 2015. "Semiannual Monetary Policy Report to the Congress." Testimony before the Committee on Financial Services, U.S. House of Representatives, July 15. http://www.federalreserve.gov/newsevents/testimony/yellen20150715a.htm

Comments and Discussion

COMMENT BY

KATHARINE ABRAHAM I take away two main conclusions from this very useful paper. First, the authors have convinced me that, at least on occasion, unusual weather can cause real problems for interpreting the monthly payroll employment estimates produced by the Bureau of Labor Statistics (BLS). Second, I am also convinced that it is possible to use data on temperature, snowfall, and so on to identify the systematic effects of unusual weather on the payroll employment series and, if desired, to remove those effects from the data. My comments mainly address whether and how the approach the authors have developed might best be applied in the production of official employment statistics. Although the paper focuses primarily on the payroll employment data, as do my comments, similar issues could be raised regarding other economic time series, and I look forward to future work that explores the effects of weather on economic measurement more broadly.

The payroll employment estimates on which most data users rely are adjusted to remove the effects of normal seasonal variation in the weather along with the effects of other predictable seasonal influences. These adjustments are not intended to account for the effects of weather that is better or worse than usual for the time of year. As the paper demonstrates, the direct effects of unusual weather on employment in the affected month can be relatively large. In addition to its direct effects, unusual weather also can cause distortions in the seasonal factors used to adjust employment estimates in other months. For example, an unusually large snowstorm that depresses employment one February might lead to a lowered expectation for employment levels in the next several Februaries. If the weather were more normal the following February, employment could look stronger than it really was. The approach described in the paper removes both the direct

and the indirect effect of unusual weather from the monthly employment estimates. It would be possible, however, to use these same methods to remove the influence that unusual weather can have on seasonal adjustment factors without removing the direct effects of unusual weather on employment in the month in which it occurs. I will come back to this point.

CALENDAR EFFECTS AND WEATHER EFFECTS In reading the paper, I was struck by the parallels between the weather effects that are its subject and the calendar effects that plagued the interpretation of payroll employment data in years past. The calendar effect with the largest effects on the payroll employment series is the so-called 4-week/5-week effect. Depending on the year, there may be either a 4-week interval or a 5-week interval between the weeks in adjacent months that include the 12th of the month and are used to determine the payroll period for which employers are asked to report. The length of this interval can have an important effect on measured employment growth. In construction, to take an example of an industry where the 4-week/5-week effect can be especially important, employment tends to rise through the spring as the weather improves, meaning that the raw growth in employment from March to April is generally larger when the interval between payroll reference periods is longer. Before this was accounted for in estimation, the growth in seasonally adjusted construction employment in a year with 4 weeks between the March and April reference periods that followed years with a 5-week interval tended to look weaker than it actually was, since the seasonal expectation for the March-to-April change was heavily influenced by the larger cumulative upswing associated with a 5-week interval. Conversely, the growth in seasonally adjusted construction employment in a year with 5 weeks between the March and April reference periods could look stronger than it actually was, especially if that year followed years with a 4-week interval (Cano and others 1996).

Through the mid-1990s, discussion of the monthly employment numbers frequently included statements that were strikingly similar in tone and content to statements about the effects of weather on the numbers quoted by Boldin and Wright. "The Employment Situation: April 1995," for example, includes the following statement:

> The lack of job growth between March and April may have reflected an unusual set of circumstances. . . . The seasonal buildup in services, retail trade, and construction from March to April had been relatively large in the previous 3 years (1992–94), partly because in each case there were 5 weeks between the two collections. As a result, this year's seasonal "expectation" (which is based primarily on the prior 3 years) was relatively large. With only 4 weeks separating the surveys, however, the time period for which hiring could take place was

> reduced. All of this likely made employment in April appear weaker than it actually was. (BLS 1995)

The likelihood that having a 4-week rather than a 5-week interval between March and April had affected the data was noted in news stories at the time (for example, see Georges 1995). Payroll survey estimation procedures that removed the so-called 4-week/5-week effect from the seasonally adjusted data were introduced for most industries in 1996 and for construction in 1997.

Different calendar effects have the potential to confound the interpretation of other economic time series. It has long been recognized that flow series such as those for production, shipments, and sales may be affected by the number of working or trading days in the month or by the timing of holidays (Young 1965; Findley and others 1998). In the monthly payroll survey, hours of work tend to be lower than would otherwise be the case when there are fewer workdays during the month or when Good Friday or Labor Day falls during the survey reference period (BLS 2015). Over time, the federal statistical agencies have developed procedures to remove these sorts of calendar effects from published seasonally adjusted estimates.

The present paper proposes that procedures similar to those used to remove calendar effects could be used to remove the effects of unusual weather from published economic data series. Whether this would be a good idea depends on what purpose the adjustments statistical agencies make to economic data series should serve. One worthy goal of such adjustments is to produce series that do a better job of capturing underlying trends. A second and somewhat different goal is to produce series that are easier for statistically unsophisticated data users to understand.

With respect to the removal of calendar effects from published seasonally adjusted data, these two goals seem to me to be largely in alignment. That is, analysts are likely to prefer series from which calendar effects have been removed, and I would guess that the typical person on the street also would understand that one does not want, for example, to say employment is growing faster or slower just because the normal seasonal upswing in employment has been measured over a longer or shorter interval.

With respect to the removal of weather effects from published seasonally adjusted data, however, the goal of producing a series that better captures an underlying trend may lead to a different conclusion than the goal of producing a series that is easier for statistically unsophisticated data users to understand. Imagine a situation in which a large blizzard had shut down economic activity across much of the country for an extended

period of time. Analysts might find an employment estimate from which the effects of that blizzard have been removed to be more useful as an indicator of underlying trends. It is difficult, however, for me to imagine the commissioner of labor statistics reporting such an estimate to the public as the official measure of what had happened to employment during the month. A number that represented what would have happened to employment if there had been no blizzard undoubtedly would be of analytical interest, but it would lack face validity as a representation of reality. For that reason, although I would value changes to its procedures that allowed the BLS to remove the distortions to seasonal factors potentially associated with unusual weather and also to better quantify the direct effects of weather on published employment estimates, I would be uncomfortable with incorporating weather adjustments of the sort described in the paper into the featured payroll employment figures.

HOW THE BLS HANDLES WEATHER ADJUSTMENT As background for thinking about how the BLS might apply the methods developed by Boldin and Wright to improve monthly payroll employment estimates, it may be useful to say a little bit about how unusual weather is handled by current BLS seasonal adjustment procedures. Seasonal adjustment of the payroll employment data is implemented by producing seasonally adjusted estimates for detailed estimation cells and then summing the resulting numbers to create seasonally adjusted employment estimates for more aggregated industries and for the nonfarm business sector as a whole. As already mentioned, current BLS procedures are not designed to account directly for the effects of unusual weather, but an estimate for a particular estimation cell that is deemed to be an outlier—as might be the case if unusually good or unusually bad weather had an especially large effect on the number for the estimation cell—may be excluded for the purpose of calculating seasonal factors. Outside the construction industry, however, this rarely happens.

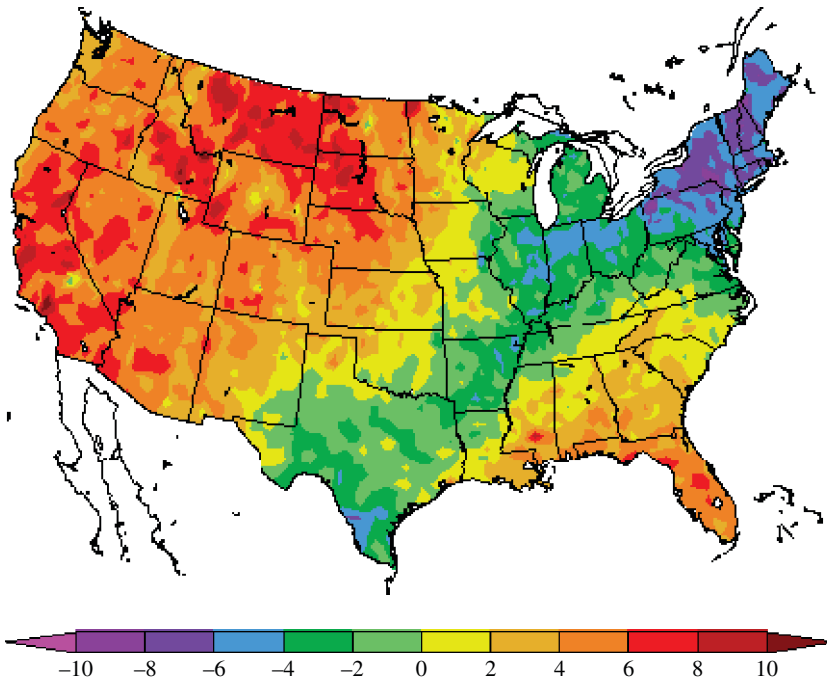
Special procedures to address the effects of unusual weather on construction employment have been in place since 1997 (Kropf 1996; Getz 1997). One year earlier, in 1996, new procedures to address the 4-week/5-week calendar effect in the payroll employment data had been introduced. Because the effects of weather on construction employment are so large, however, usable 4-week/5-week adjustment factors could not be estimated for construction without taking weather effects into account, and the implementation of the new 4-week/5-week procedures in construction had to be delayed. This made it a priority to develop some method for addressing the effects of weather on construction employment.

Within construction, payroll employment estimation, including seasonal adjustment, is carried out at the most detailed industry level for which data are available—either the 5-digit or 6-digit North American Industry Classification System level—and, where possible, separately for each of four regions. Estimates for the relevant detailed industry cells or the detailed industry by region cells then are summed to produce national estimates for published industries. Within construction, the bounds used to determine whether a monthly estimate is an outlier are set to be tighter so that estimates are more likely to fall outside the defined bounds and be classified as outliers. Analysts verify apparent outliers in the construction employment estimates as weather-related by checking against information from the National Weather Service and then, if appropriate, they remove the outliers from the data series used to calculate seasonal factors.

As a historical footnote regarding the approach the BLS has adopted to deal with the effects of unusual weather on construction employment, I have been told that when developing its special procedures for construction, the BLS asked the National Weather Service for data on average temperature to use in estimating the effects of weather but was turned down. As I understand it, the National Weather Service explained that weather conditions can vary considerably across different parts of the country and information on average temperature would be meaningless. Boldin and Wright make a good case that measures of average weather could in fact have been very useful! That said, recognizing that there is variation in weather conditions across different parts of the country could allow the BLS to improve on Boldin and Wright's suggested method of accounting for weather effects.

THE CHALLENGE OF GEOGRAPHIC VARIATION Consider the effects of temperature on employment. The measure employed in the analysis reported in the paper is a measure of the average across weather stations of the deviation of temperature from its normal level at that weather station in a given month. In many months, however, conditions may be unusually hot in some areas but unusually cold in others. As an illustration, my figure 1, a chart prepared by the High Plains Regional Climate Center and disseminated by the National Weather Service, displays the deviations of the average temperatures from their historical mean levels in different areas for March 2015. Temperatures were considerably below average that month in the Northeast but considerably above average in the Southeast and West. Similar variation may be observed in the monthly data for snowfall, precipitation, and so on.

Figure 1. Departure from Normal Temperature, March 2015[a]



Source: High Plains Regional Climate Center (http://www.hprcc.unl.edu/products/maps/acis/Mar15TDeptUS.png).
a. Temperature measured in degrees Fahrenheit.

This sort of variation would not matter if the effects of deviations from normal weather conditions were both linear and of the same magnitude in all locations. This is unlikely to be the case. The effect of being above or below average with respect to temperature, snow, or other weather conditions in a month can vary substantially by region. Weather that was 10 degrees warmer than usual during February, for example, could have a significant effect on employment in Boston but no effect on employment in Phoenix. This implies that a warmer-than-usual February might or might not be associated with higher-than-average employment, depending on where the warmer-than-usual weather occurred. Similarly, an extra six inches of snow might have no effect on employment if it falls in Minneapolis, but a disastrous impact on employment if it falls in Atlanta. Again, in a month in which average snowfall was greater than expected, it would matter where the extra snow had fallen. The fact that Boldin and Wright obtain better model fits with their preferred snow variable—constructed as

the weighted average of regional measures of the societal effects of different storms rather than average snowfall—is consistent with the idea that deviations of weather from its norm may have different effects in different regions. Thus, important information is lost by relying on national average weather measurements to make the weather adjustments.

As explained in the paper, it would not have been possible for Boldin and Wright to implement a geographically disaggregated weather adjustment using published BLS data. The published national employment series refers to the country as a whole, and the state-level employment estimates that the BLS also publishes do not sum to the national estimates. Internally, however, the BLS already makes use of regionally disaggregated estimation cells for construction employment, and conceivably it could do the same for other weather-sensitive industries. This means that, at least in construction, the BLS already has a natural platform in place for incorporating regional weather information into its estimation procedures. My guess is that weather adjustments based on regional weather data might be at least somewhat larger in size than those reported by Boldin and Wright, though this is of course an empirical question.

THE CHALLENGE OF PAYROLL VARIATION AMONG INDUSTRIES There is one other respect in which the methods outlined by Boldin and Wright might be improved upon. As explained in the paper, the weather variables used for adjusting the employment data are created by weighting weather measurements for the 30 days prior to the 12th of the month, with the coefficients of the parametric function used to define the relative weights accorded to different days selected to maximize the fit with national employment data. The important point is that these relative weights are restricted to be the same across all industries. It seems plausible, however, that the relative importance of weather on different days prior to the 12th could vary across industries. It might matter, for example, whether work in the industry is done inside or outside, whether employees in the industry are able to work remotely, and whether and how weather affects the demand for the industry's products or services. I suspect that improving the weights accorded to weather on the different days in the month before the 12th is a second-order issue, but it might nonetheless be worth investigating.

CONCLUSION Supposing that the BLS were to decide to adopt the methods developed by Boldin and Wright—something that I think is very much worth considering—there is still the question of exactly how they would be used. One obvious application would be to use Boldin and Wright's methods in developing seasonal adjustment factors for the official payroll employment statistics that are not contaminated by the effects of unusual

weather. I also would like the BLS to report the estimated magnitude of the effects of weather on each month's employment and perhaps even to prepare research or supplemental series from which weather effects have been removed. From my perspective, however, to the extent that weather affects the level of employment in a particular month, that should be reflected in the official payroll employment numbers.

REFERENCES FOR THE ABRAHAM COMMENT

BLS (Bureau of Labor Statistics). 1995. "The Employment Situation: April 1995." News Release. Washington. http://www.bls.gov/news.release/history/empsit_050595.txt

———. 2015. "Technical Notes for the Current Employment Statistics Survey." Washington. http://www.bls.gov/web/empsit/cestn.pdf

Cano, Stephanie, Patricia Getz, Jurgen Kropf, Stuart Scott, and George Stamas. 1996. "Adjusting for a Calendar Effect in Employment Time Series." In *Proceedings of the Survey Research Methods Section.* Alexandria: American Statistical Association.

Findley, David F., Brian C. Monsell, William R. Bell, Mark C. Otto, and Bor-Chung Chen. 1998. "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program." *Journal of Business and Economic Statistics* 16, no. 2: 127–52.

Georges, Christopher. 1995. "Economists Say Chances of a Recession Remain Slim, in Spite of Recent Data." *Wall Street Journal*, May 8.

Getz, Patricia. 1997. "Improved Seasonal Adjustment for Construction." Unpublished memorandum for George Werking, April 15. Washington: U.S. Bureau of Labor Statistics.

Kropf, Jurgen. 1996. "4/5 Adjustment for Construction, SIC 2015, 2016 and 2017." Unpublished memorandum for Pat Getz, August 21. Washington: U.S. Bureau of Labor Statistics.

Young, Allan. 1965. "Estimating Trading-Day Variation in Monthly Economic Time Series." Technical Paper no. 12. Washington: U.S. Bureau of the Census.

COMMENT BY

CLAUDIA SAHM[1] Michael Boldin and Jonathan Wright introduce a new method for estimating the impact of weather on key economic data series, like monthly payroll employment. Their aim is to provide a

1. I am thankful to Steve Braun, Tyler Cowen, Charles Gilbert, Norman Morin, and Andrew Paciorek for helpful conversations that informed my comments, and to Erik Larsson for his great research assistance. These are my views and are not necessarily shared by others in the Federal Reserve System or the U.S. government.

clearer view of business-cycle fluctuations by removing weather effects. The authors extend a widely used seasonal adjustment algorithm, which already isolates calendar effects and with it the "usual" weather changes over the year. They add a first-stage estimation to the algorithm with a direct measure of weather, so their extended algorithm isolates the impact of both usual and unusual weather.

Trying to estimate the impact of unusual weather events on economic data has a long history among macroeconomic forecasters, so the contribution of this paper is a technical improvement: examining a large set of weather measures, using disaggregated industry data, and working within the existing seasonal adjustment framework. While there is more work to be done, this analysis could serve as the basis for systematic weather adjustment in official statistics.

This new seasonal-and-weather adjustment algorithm would be particularly useful to individuals who need to interpret economic conditions in real time. The difference between slow demand due to severe weather and slow demand due to an incipient recession is crucial to many economic decisionmakers, including central bank officials setting interest rate policy and business managers weighing new investments. In fact, the importance of isolating weather effects is borne out by the cottage industry of macroeconomic forecasters who have provided such estimates for years.

Nonetheless, the winter of 2014 provides a good example of how this paper can add value. During that period, the country experienced one set of weather conditions and one realization of economic activity (though the latter did revise over time), and yet there was a wide range of professional estimates on the output effect from the severe weather. Macroeconomic Advisers (2014) at the time estimated that "elevated snowfall . . . reduced first-quarter GDP growth by 1.4 percentage points," while an analysis from Goldman Sachs maintained that "weather [would] cause first-quarter GDP to be 0.5 [percentage] point worse than it otherwise would have been" (Goldstein 2014). Federal Reserve staff characterized the weak GDP data this way: "Unusually severe winter weather could account for some, but not all, of the recent unanticipated weakness" (FOMC 2014).

And while few decisions hinge on the exact estimates of weather effects, the extent to which a shift in economic activity can be explained by weather is important. That is because the weather events considered by the authors, such as a severe winter storm, are viewed as a temporary shock and something that most economic decisionmakers should see through. A snowstorm may keep a consumer from buying a car at the end of January, but presumably when the weather clears, she will still buy the car. That kind of

short-term delay—shifting output from one month to the next—should not concern policymakers, though a drop in car purchases due to diminished job prospects would. In real time, when one does not yet know the next month's or next quarter's data, the source for a drop in spending can be difficult to determine.

I applaud the authors' efforts to bring more technical discipline to estimating weather effects, yet I have three concerns with the paper. First, I think there needs to be more discussion about the relative importance of unusual weather and the danger of elevating this transitory shock simply because it is something visceral. Business-cycle fluctuations will always be somewhat obscured by noise in the data. Second, I think the authors need to do more to develop the diagnostics of the algorithm. There needs to be clearer guidance on when to use their seasonal-and-weather adjustment similar to the guidance from statistical agencies on when to use the standard seasonal adjustment. And third, in making inferences about weather's impact, one needs to explore how the weather impact may depend on the business-cycle conditions. I am concerned that this research brings us from removing usual winter weather in usual business-cycle conditions to removing both unusual and usual weather in usual business-cycle conditions. This is a step forward, but it may not fully capture how much a particular month's or quarter's data are affected by a weather event. Before we begin filtering all our economic data with this new algorithm, we need to think more about the counterfactual—what the world would have looked like without the weather event—and the variation we would be removing from our economic analysis.

High-frequency economic data can be quite noisy. For example, the 90-percent confidence interval on the monthly change in total nonfarm payroll employment is plus or minus 115,000.[2] Many of the weather effects that the authors highlight, such as the 64,000 reduction in payroll employment in February 2014, are well within the confidence intervals that reflect sampling and nonsampling error. Still, unusual weather occurs often enough, and comments from Federal Open Market Committee (FOMC) minutes in the month of March for recent years show that we need a reliable method for isolating such weather effects (see the first column in my table 1). Nonetheless, seasonally-and-weather adjusted data should not give us a false sense of clarity. The second column of my table 1 shows other, non-weather events that were mentioned in the same FOMC minutes

2. See Bureau of Labor Statistics, "Employment Situation Technical Note," February 5, 2016, http://www.bls.gov/news.release/empsit.tn.htm.

Table 1. Events Noted as Obscuring Underlying Economic Conditions in March FOMC Minutes, 2010–15

	Weather	*Non-weather*
2015	"unseasonably cold winter weather"	"labor disputes at West Coast ports"
2014	"unusually cold and snowy winter weather"	"partial government shutdown"
2013	—	"federal spending sequestration"
2012	"unseasonably warm weather"	—
2011	"weather-related distortions in various indicators"	"earthquake, tsunami"
2010	"adverse effects of the snowstorms"	"waning effects of fiscal stimulus"

Sources: FOMC (2010, 2011, 2012, 2013, 2014, 2015).

as also obscuring underlying economic conditions. The regularity of some shocks early in the year, such as discretionary changes in fiscal policy, also caution against writing off all the recent first-quarter weakness in recent years as an inability to remove calendar or weather effects.

The authors weather-adjust all the data series regardless of how well the weather model fits an industry series, but this decision is at odds with the standard use of seasonal adjustment in official statistics. For example, consider this statement from the U.S. Census Bureau:

> The Census Bureau performs seasonal adjustment of a time series of estimates only given clear evidence of seasonal behavior and only when the adjustment passes a suitable set of diagnostic tests. (McDonald-Johnson and others 2010)

Charles Gilbert and others (2015) provide an example of using such diagnostic tests to examine residual seasonality in output data. Stability of the adjustment factors is a guiding principle for the decision as to when it is appropriate to seasonally adjust a data series. One might view the diagnostic tests for seasonal adjustment in official series as too stringent, but there needs to be further analysis of how stable the seasonal-and-weather adjustments in the Boldin and Wright paper are. The stability of the weather impact estimates may be improved by focusing on series that show a clear weather impact, such as construction employment.

Finally, it is important to take a step back and think about the variation being removed with the seasonal-and-weather adjustment. Consumers, employers, and even policymakers experience the economy with all its seasonal and weather-related variation, so using adjusted data misses the opportunity to study that variation. Robert Barsky and Jeffrey Miron (1989) argue, for example, that seasonal variation could be used to test macroeconomic models, yet macroeconomic studies with not seasonally adjusted data are exceedingly rare. Of related concern, the weather impact

may not be neatly separable from underlying economic conditions. Alan Auerbach and Yuriy Gorodnichenko (2012) argue that the impact of discretionary government spending on output (the fiscal multiplier) is larger in recessions than in expansions. Likewise—and a point acknowledged but not explored by the authors—a severe snowstorm may have a different impact on activity during a recession than during an expansion. This would complicate the full removal of weather effects, and the removal, even partially, may sacrifice some information on underlying economic conditions.

As a simple example of how weather might interact with the business cycle conditions, I estimated a standard model of monthly retail sales growth (*RS Growth*), which includes heating degree days (*HDD*), consumer sentiment (*Sent*), and an interaction between them. I chose sentiment as a business cycle indicator because, unlike some employment series, it does not vary with the weather measure in the regression. The results of the estimation are represented below, with standard errors in parentheses.[3]

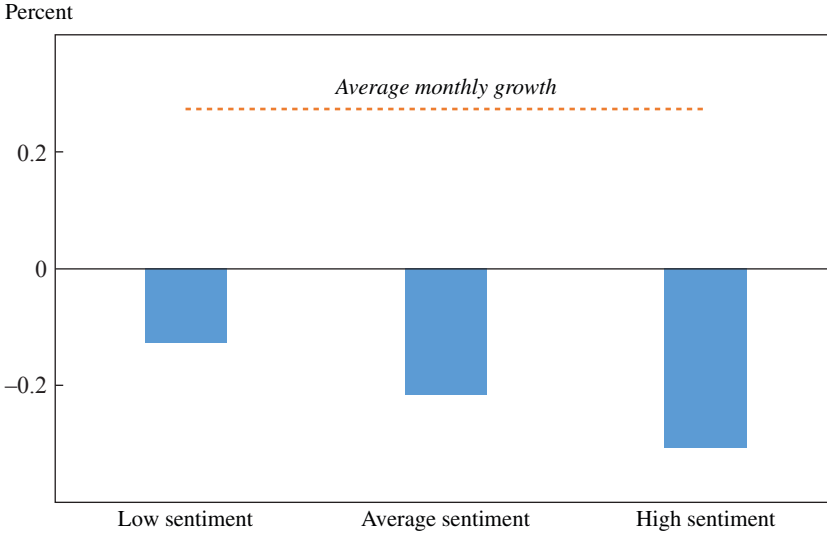
$$RS\ Growth_t = \underset{(0.045)}{0.28} - \underset{(0.047)}{0.25}HDD_t + \underset{(0.047)}{0.14}HDD_{t-1} + \underset{(0.042)}{0.10}Sent_t - \underset{(0.049)}{0.09}HDD_t \times Sent_t$$

Unusually cold weather, which is a positive heating degree day reading, depresses the growth in retail spending in the current month and boosts it in the subsequent month, highlighting the transitory nature of weather shocks. This is a well-known feature of retail sales growth—an example of how weather estimates are often done. The positive association between retail sales growth and sentiment is also standard. The additional feature of this simple model, as shown in my figure 1, is that unusually cold weather weighs more on retail spending growth at times when sentiment is high, measured as one standard deviation above average.

Intuitively, it makes sense that if economic activity is picking up and consumer sentiment is high, a severe winter storm would imply a large drag on growth, since there is more growth to disrupt relative to the counter-

3. The regression is estimated with monthly data from January 1999 to July 2015, and the R^2 is 0.17. All of the coefficients are statistically significant at the 10 percent level. The dependent variable is seasonally adjusted retail sales excluding autos, gasoline, and building materials; this is the portion of the retail sales data used by the Bureau of Economic Analysis in its estimate of personal consumption expenditures, and it accounts for roughly one-fifth of GDP.

Figure 1. Impact of Severe Cold on Retail Sales Growth by Business Cycle Conditions



factual world of normal weather. In isolating the impact of weather, it is better to remove the "average sentiment" and heating degree effect, which is closest in spirit to the authors' seasonal and weather-adjusted data but that does not mean that all of the weather impact has been removed. And this also leaves open an interesting question about how consumers or employers interpret these weather shocks, which in real time would be hard to distinguish from other economic shocks. With its careful technical treatment of estimating weather effects, this paper should serve as an invitation to think more about what weather's impact is on the economy.

REFERENCES FOR THE SAHM COMMENT

Auerbach, Alan J., and Yuriy Gorodnichenko. 2012. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Economic Policy* 4, no. 2: 1–27.

Barsky, Robert B., and Jeffrey A. Miron. 1989. "The Seasonal Cycle and the Business Cycle." *Journal of Political Economy* 97, no. 3: 503–34.

FOMC (Federal Open Market Committee). 2010. "Minutes of the Federal Open Market Committee: March 16, 2010." Washington: Board of Governors of the Federal Reserve System. http://www.federalreserve.gov/monetarypolicy/fomcminutes20100316.htm

———. 2011. "Minutes of the Federal Open Market Committee: March 15, 2011." Washington: Board of Governors of the Federal Reserve System. http://www.federalreserve.gov/monetarypolicy/fomcminutes20110315.htm

———. 2012. "Minutes of the Federal Open Market Committee: March 13, 2012." Washington: Board of Governors of the Federal Reserve System. http://www.federalreserve.gov/monetarypolicy/fomcminutes20120313.htm

———. 2013. "Minutes of the Federal Open Market Committee: March 19–20, 2013." Washington: Board of Governors of the Federal Reserve System. http://www.federalreserve.gov/monetarypolicy/fomcminutes20130320.htm

———. 2014. "Minutes of the Federal Open Market Committee: March 18–19, 2014." Washington: Board of Governors of the Federal Reserve System. http://www.federalreserve.gov/monetarypolicy/fomcminutes20140319.htm

———. 2015. "Minutes of the Federal Open Market Committee: March 17–18, 2015." Washington: Board of Governors of the Federal Reserve System. http://www.federalreserve.gov/monetarypolicy/fomcminutes20150318.htm

Gilbert, Charles E., Norman J. Morin, Andrew D. Paciorek, and Claudia R. Sahm. 2015. "Residual Seasonality in GDP." *FEDS Notes*. Washington: Board of Governors of the Federal Reserve System.

Goldstein, Steve. 2014. "March Was Colder but Less Snowy Than Usual, So Data Should Get Lift, Forecaster Says." Blog post, April 1. MarketWatch.

Macroeconomic Advisers. 2014. "Elevated Snowfall Reduced Q1 GDP Growth 1.4 Percentage Points." Blog post, April 16. http://www.macroadvisers.com/2014/04/elevated-snowfall-reduced-q1-gdp-growth-1-4-percentage-points/

McDonald-Johnson, Kathleen M., Brian Monsell, Ryan Fescina, Roxanne Feldpausch, Catherine C. Harvill Hood, and Monica Wroblewski. 2010. "Seasonal Adjustment Diagnostics." Census Bureau Guideline, Version 1.1. Washington: U.S. Census Bureau. https://www.census.gov/ts/papers/G18-0_v1.1_Seasonal_Adjustment.pdf

GENERAL DISCUSSION Jonathan Pingle wondered if Michael Boldin and Jonathan Wright had checked the stability of the coefficients that they estimated in their model. Pingle noted that, in his own work, the impact of snowfall appeared to have changed over the course of the past three decades.

Pingle also wondered if there is evidence of asymmetry in the data; that is, does the weather being better-than-normal by a certain amount have an equal and opposite effect as the weather being worse than normal by the same amount? As an example, he noted that it is often the case where it seems like a very cold March will be followed by instant bounce-back in April, but a very warm March sometimes seems to pull forward seasonal inflows for several months. He was curious if that kind of asymmetry might be driving some of the lagged effects mentioned in the paper.

Along similar lines, David Romer wondered about the assumption often made by short-term forecasters of full bounce-back due to the effects of weather; on average, that is, are the negative effects of bad weather made up the following month as the weather on average returns to normal? For example, if weather reduced employment in one month by 50,000 jobs, should it really be assumed that the next month is going to add those 50,000 jobs? Perhaps that would make sense, he noted, in a world where firms have a set number of people they want to hire, and they can just go out and get them. But perhaps matches that do not occur in one month are not magically formed in the next.

Pingle wondered if Boldin and Wright had given any thought to the implications of the difference observed between the weather effects in the initial release of the Current Employment Statistics (CES) data and the weather effects in the revised data. Sometimes, the initial release seems to lack a significant weather effect, but it is more pronounced upon revision. He wondered if it could be the case that more weather-affected establishments were not reporting in as timely a manner. If that were the case, then applying Boldin and Wright's methodology to the initial release of the data could offset a negative weather effect not yet in the data, thus overstating the month's employment.

Jeff Campbell wondered about the implications of the model as it relates to forecasting. He agreed with discussant Katharine Abraham, who had noted that one of the key goals of weather adjustment might be to produce data series more suitable for short-term forecasting, noting that there are alternative means of doing that. In the methodology implemented by Boldin and Wright, weather adjustment is applied to an economic series before being put into a forecasting model. Campbell wondered if a principal component of the weather adjustment could be applied to the forecasting model itself. Alan Blinder took issue with the claim that a goal of adjusting data in the first place is to make them more suitable for short-term forecasters, noting that a broader group of professionals is interested in adjusted data. Campbell also wondered if authors had any results relating to inventories.

Abraham also had suggested that another goal of adjusting data might be to produce series that are easier for ordinary people to understand. Blinder took strong issue with this suggestion as well. He noted that adjusted data are used by only a small cadre of experts, and that ordinary people live in a real world that is not seasonally or weather adjusted. For example, if an ordinary person wanted to get a job in retail, it would be easier for her to look during the Christmas season rather than in January; if she wanted to

be a lifeguard, she would have an easier time finding a job in early summer rather than in October. For the small cadre of experts, however, these adjustments are really useful.

Valerie Ramey, who had dealt extensively with weather effects in her work, remarked on how useful the Boldin and Wright methodology was. She appreciated Boldin and Wright's systematic approach to removing large weather-related outliers from the data. Ramey suggested that some of the assumptions made by Boldin and Wright were appropriate, while others might need to be loosened. She agreed that the effects of lags considered by Boldin and Wright were definitely necessary, citing her experience with auto assembly plants. She recounted the great blizzard of 1978, in which assembly line workers were able to return to work, but the plants remained closed because the blizzard had prevented the delivery of parts.

Ramey took issue with the authors' model identification in two aspects. First, she questioned the assumption that weather occurring after the 12th day of the month should not have any effect on employment data. In most surveys that measure employment data, the reference period is generally the calendar week or pay period that contains the 12th day of the month. She suggested that the authors consider at least a few days after the 12th, noting that forecasts of abnormal weather in the near future may actually affect employment in the present. Steve Braun echoed Ramey's concerns, noting that if the 12th occurred on a Sunday, then as many as five extra days might need to be considered.

Ramey also questioned the assumption that unusual weather events do not have permanent effects, citing a growing environmental literature. She noted, as examples, that hurricanes hitting small islands, or Hurricane Katrina hitting New Orleans, certainly had some permanent effects on employment.

Robert Gordon commented on the relationship between GDP and payroll employment, namely productivity. Payroll data do not reflect big weather events as strongly as data for GDP, resulting in overinflated estimates of productivity and extremely high positive correlations between output and productivity. Looking at detrended levels of output productivity over the last five to twenty years, Gordon noted it is clear that there is no longer any short-term positive correlation between productivity and output. He posited that the historically low productivity growth data over the last five years reflect structural rather than cyclical changes in the economy.

Braun praised Boldin and Wright's methods as clear improvements over previous efforts to model weather adjustment. He suggested that it might be useful to apply the model not only to employment, as Boldin and Wright

do, but also to measures such as work week and man-hours, which are more sensitive to the effects of weather. He noted that the exercise might be particularly interesting because the correlation between man-hours and GDP is much stronger than between employment and GDP, a concern raised earlier by Gordon.

Christopher Carroll pointed out that there may be substantively important issues on which forecasters could reach the wrong conclusions about the underlying momentum of the economy if they did not take into account some kind of weather adjustment in their forecasts. According to Boldin and Wright, the serial correlation of growth is substantially greater when adjusted for weather. Historically, the debate about whether or not the serial correlation of a variable was important centered on whether or not consumption followed a random walk. Carroll believed that evidence seemed overwhelming that there was a lot more momentum in consumption growth than was apparent originally because weather effects were adding noise to quarterly numbers. He was enthusiastic about the work put forward by Boldin and Wright, speculating that it might have real consequences for how business cycle models are calibrated.

Andrew Abel commented on the differences between the Current Population Survey (CPS) and the CES in how they measure employment. In the CES, abnormal weather that prevents people from going to work reduces the payroll employment count, since the data are based on surveys of businesses. In the CPS, however, an individual who does not go to work due to abnormal weather may still be counted as employed if weather or some other reason is given for not working. Abel wondered if Boldin and Wright were aware of how many people responded to the CPS in this way, and whether the magnitudes compared with the magnitudes that the authors calculated with their weather-adjustment model.

Adele Morris was interested in understanding regional and local labor market vulnerabilities to extreme weather events, and promoted Abraham's suggestion to have regional data adjustments in addition to the national adjustments made by Boldin and Wright. If more extreme weather events were on the horizon, it would be beneficial to have a deeper understanding of the regional and local vulnerabilities. In addition to trying to take out weather information to see what remains, she suggested that what is taken out may be extremely interesting to people who are thinking about local or regional policies to adapt to a changing climate.

Discussant Claudia Sahm brought up the question of when it is appropriate to adjust for weather, and which series to adjust. One could argue that the Census is too conservative in deciding which series to adjust, that is,

it should be adjusting more series than it currently does. When adjusting a series for weather affects, it is important that the effect pulled out is in fact a weather effect, and not some other kind of effect. It is clear that employment data in construction should be adjusted because there are clear and stable patterns. However, other series might not benefit from weather adjustment, since it might not be clear that they are stable or significant enough.

Sahm believed that looking at disaggregated data could be important, for instance in analyzing the behaviors of consumer spending. In the case of motor vehicle consumption, bad weather may deter the purchase of a vehicle in one month, but the vehicle would almost certainly be purchased in the next month. On the other hand, if a consumer was prevented from going out to dinner or purchasing something for the holidays, persistent bad weather may actually prevent those transactions from ever happening in the near future.

Responding to Pingle's and Romer's question of asymmetry, Wright noted that he and Boldin looked into the issue, but did not find much evidence. The authors described an experiment in which they considered weather in the previous month and weather in the previous 2 months; they found that bad weather one month prior did lower the level of employment, but 2 months prior did not. He conceded that it is not correct to say that there is complete bounce-back immediately, but that it is not very far off.

In response to Ramey's and Braun's concerns about abnormal weather after the 12th day of the month, Wright stated that he and Boldin did test the effects of weather after the 12th day of the month, but only in a very crude way; by adding extra variables for the weather on the 13th and 14th, Boldin and Wright found that they were insignificant in the aggregate. Wright agreed with Ramey that some weather events may have a permanent impact, but argued that her framework might not be the right way to think about what the model was meant to isolate, such as a snowy winter or colder-than-normal January.

Wright responded to the question raised by Sahm regarding which series to adjust. If the ultimate interest is in aggregates rather than disaggregates, then Wright believes that statistical agencies are currently too conservative and too willing to just decide not to seasonally adjust a series at all.

Wright noted that while technology for weather adjustment is available and hopefully useful, it is not intended to replace the data. It is useful to have some way to figure out what the effects of weather are, and for statistical agencies to have some advantage in being able to do that.

JULIAN SCHUMACHER
Johannes Gutenberg University Mainz

BEATRICE WEDER DI MAURO
*INSEAD**

Greek Debt Sustainability and Official Crisis Lending

ABSTRACT The International Monetary Fund and the European Stability Mechanism softened their crisis lending policies repeatedly to deal with the Greek debt crisis, but the analysis of debt sustainability still acts as the gatekeeper for access to official financing. We explore the underlying mechanics of debt sustainability analysis and show that the standard model is inappropriate for Greece since it ignores the highly concessional terms of Greek debt. Greek debt has been restructured repeatedly, and now two-thirds of the stock contains grant elements of about 54 percent. The present value of outstanding Greek debt is currently about 100 percent of GDP and will rise to about 120 percent under the new program. Greek debt sustainability therefore is less a problem of the debt stock. By simulating different paths of the gross financing needs, we show that there may be liquidity problems over the medium to long terms (in particular, in 2035 and beyond). However, our estimation of the financing need is subject to high uncertainty and mainly depends on whether Greece will be able to regain access to markets at reasonable terms.

Many people hold strong views on Greek debt. Just taking a casual look at the level of Greece's debt, which the International Monetary Fund (IMF) has recently projected to rise to 200 percent of GDP, it seems obvious that Greek public finances cannot possibly be sustainable (IMF 2015a, 2015b). What is not obvious is how this can be subject to dispute among the main creditor institutions. Greek official sector debt sustainability assessments have been quite volatile, but by the beginning of 2015 the verdict of the main official creditor institutions—hereafter referred to as the *troika* for short—was that Greek debt was sustainable.[1] Eight months

*The author is currently on leave from Johannes Gutenberg University Mainz.

1. The troika consists of the European Central Bank, the European Commission, and the International Monetary Fund.

later, however, at the time of this writing, the troika has split over the question of debt sustainability; while the Europeans are pushing ahead with a new program for Greece, the IMF is holding out. It seems that the diagnosis of debt sustainability is not so obvious after all.

One reason that a diagnosis of debt sustainability is complicated could be that politics plays a role, in particular the political feasibility of adjustment. Indeed, theory has long emphasized that sovereign debt is different from corporate debt, precisely because politics and institutions are crucial in determining a country's capacity *and* willingness to repay.[2] From this perspective, debt sustainability would depend, among other things, on the particular political coalition, the strength of political institutions, and even on the egos of decisionmakers and their negotiating power, both at home and abroad. Thus, debt sustainability would not only be unobservable and country-specific but also time-varying and highly volatile. Political positions can change very rapidly, as showcased by the turmoil caused by the Greek government's turnover in January 2015.

However, this is a perspective that neither the IMF nor the eurozone can adopt. As a matter of principle, they must ensure equal treatment across members and cannot constantly change the goalposts in accordance with shifting political circumstances. Thus, they need to deploy a framework to assess debt sustainability that can be applied to the entire membership.[3] In addition, this framework should be designed with the goals of both protecting the debtor country from overborrowing and protecting the resources of the creditor institution. An "unsustainable" verdict should preclude the official sector from lending into cases of insolvency and should require some form of debt restructuring first (IMF 2014a, 2014c). Understanding the basis of official debt sustainability analysis is therefore crucial. The first contribution of this paper (section I) is to present the models for sustainability assessment employed by the IMF and the European Stability Mechanism (ESM), and then to review the impact of the Greek debt crisis on the overall framework for international crisis lending.

2. A large body of theoretical literature emphasizes this point, starting with Eaton and Gersovitz (1981) and summarized extensively in the surveys by Eaton and Fernandez (1995); Panizza, Sturzenegger, and Zettelmeyer (2009); and Aguiar and Amador (2014). The empirical literature has identified many political and institutional factors that have an impact on the probability of sovereign default; for example, see Van Rijckeghem and Weder (2009), Tomz and Wright (2013), and Sandleris (2015).

3. In its analytical framework, the IMF (2013b, p. 4) does make a reference to the political feasibility of primary balance adjustments, but this does not seem to depend on the country-specific political situation.

The paper's second contribution (section II) is to evaluate present Greek debt sustainability in light of these models. We show that neither of the current frameworks sufficiently takes into account the extent of Greece's dependence on official sector funding. We analyze Greek debt using a present-value approach to account for the concessionality element. Our main finding is that the overall debt contains a grant element of 37 percent and the European loans of up to 60 percent. We also show an analysis of projected repayment flows, or gross financing needs. While the main projection shows a critical level of more than 15 percent over the long term, this result is highly sensitive to assumptions about market conditions in the coming decades.

We are not the first to argue that the face value of (gross) sovereign debt may be a misleading measure. For rich but highly indebted countries like Japan, it has long been suggested that government debt should be measured in net rather than gross terms by deducting the value of government assets from the debt stock (IMF 2013b). A more recent suggestion is that sovereign debt should be expressed according to international accounting standards, including measuring the debt stock at fair values (Serafeim 2015; Kazarian 2015). We are more in line with Daniel Dias, Christine Richmond, and Mark Wright (2014), who show that measuring debt in present value, rather than face value, enables cross-country comparisons and discourages the hiding of true indebtedness behind convenient debt profiles. But our main point applies to Greece and countries borrowing from the ESM. For them, looking at gross debt will lead to a misdiagnosis, because it does not appreciate the concessional nature of European crisis lending.

The remainder of the paper is organized as follows. In section I, we discuss the role of debt sustainability analysis within the European and IMF crisis-lending frameworks. In section II, we highlight the uncertainty of Greek debt sustainability. And in section III, we provide our policy conclusions.

I. The Official Sector's Lending Framework and the Role of Debt Sustainability Analysis

The lending frameworks of the IMF and the ESM set the parameters under which countries in financial distress are considered illiquid, rather than insolvent, and can therefore receive emergency financing without first restructuring existing debt. The principle of committing to not lend into unsustainable debt dynamics is justified by several reasons, most importantly by the need to protect taxpayer resources and prevent the debtor from

accumulating excess debt, which would make future adjustment and eventual restructuring more costly. Nevertheless, this commitment is regularly tested, since the official sector will be tempted to lend even in highly doubtful cases, hoping that conditions will improve, and thus avoid the immediate costs of a restructuring, a policy known as "kicking the can down the road." The temptation to lend into insolvency is even larger if there are fears of global or regional contagion resulting from a restructuring.

Consequently, debt restructurings tend to do too little, come too late, and are too costly (Levy Yeyati and Panizza 2011; IMF 2013a). Moreover, such time-inconsistent policies of official lenders may result in overborrowing. The presence of an international lender of last resort creates incentives for private creditors to lend without regard for risk in the expectation of an official bailout. The costs of overborrowing and delayed restructuring are then mostly borne by local taxpayers, since official lenders tend to be repaid (Buchheit and others 2013). By governing the decision to provide emergency funds or insist on debt restructuring and relief first, the lending frameworks of the IMF and ESM in practice act as sovereign debt restructuring regimes. The analysis of debt sustainability is their main gatekeeper.

I.A. The ESM and IMF Crisis Lending Frameworks

The European crisis lending framework, as laid out in the 2012 treaty establishing the ESM, provides for rule-based decisionmaking for the granting of emergency loans. Article 13 requests that an application by a member state will be considered based on an assessment of three criteria through the European Commission (EC) in conjunction with the European Central Bank (ECB): (i) The risks to the financial stability of the euro area as a whole; (ii) the sustainability of public debt (if appropriate, in conjunction with the IMF); and (iii) the actual or potential financing needs of the applicant member state.

In principle, ESM loans will only be extended if the member state's public debt is sustainable. However, the treaty does not give clear guidance on how to proceed if the results of the ESM-EC-ECB debt sustainability analysis indicate an unsustainable situation.[4] Specifically, there are no provisions that would require a debt restructuring to unlock ESM access in a case where the sustainability analysis suggests an unsustainable debt.

4. In the following, we refer to the debt sustainability analysis framework mentioned in the European Stability Mechanism treaty, which is conducted jointly by the European Commission and European Central Bank, as the "ESM framework."

The IMF's lending "of last resort" to countries in financial trouble is based on multiple variables. First and foremost, a country's maximum loan volume is determined by its "quota"—a blended measure of a nation's GDP, financial openness and volatility, and official reserves (IMF 2008). In normal circumstances, countries are allowed to borrow up to 200 percent of their quota during a 12-month period, and not more than a cumulative 600 percent of their quota. The Greek quota, for instance, currently stands at SDR 1.1 billion, or about $1.5 billion, which would have limited the maximum Greek borrowing from the IMF to about $9 billion.[5] However, in exceptional circumstances, member countries are allowed to borrow more than the normal limits under the "exceptional access" policy.[6] To obtain exceptional access under the rules prevalent at the time of the first Greek program, four criteria had to be met (IMF 2004, p. 4):

i. The country is under exceptional balance-of-payments pressure exceeding the normal limits.
ii. A debt sustainability analysis indicates a high probability that the debt will remain sustainable. If the debt sustainability analysis cannot conclude this with high probability, exceptional access may be granted on grounds of systemic concerns (a "systemic exemption").
iii. The country has good chances of regaining access to private markets before the bailout ends.
iv. The country has a policy program convincingly promising success, as well as the institutional quality to implement the program.

As with the ESM framework, the IMF's framework requires an in-depth debt sustainability analysis of the country's debt stock. Before the introduction of the systemic exemption, the outcome of this analysis determined whether debt restructuring was required before a loan could be granted (IMF 2014c). Only if the debt level was deemed sustainable with high probability could exceptional access be granted without recourse to debt

5. The IMF's country quotas are expressed in terms of SDRs (Special Drawing Rights), the IMF's reserve asset based on a basket of four major currencies.

6. Access to this exceptional credit facility is determined by additional variables beyond the country quota. Providing large loans to countries in financial distress comes at greater risks, and granting exceptional access has therefore been controversial ever since Mexico received the first such loan of 688 percent of its quota in 1995 (IMF 1995). With the experience of further emerging markets crises in the 1990s in mind, the IMF formalized this instrument in the early 2000s (IMF 2002, 2004).

restructuring. The introduction of an exemption in cases of systemic concerns was therefore a major softening of the lending framework.

I.B. The Softening of the Lending Framework in Response to the Greek Crisis

European and international institutions of crisis lending were profoundly affected by the Greek crisis. In the case of the eurozone, the crisis led to the very creation of a multilateral institution for emergency financing. The previous regime only foresaw offering financial assistance for balance-of-payments crises in European Union (EU) members outside the eurozone. Inside the eurozone, fiscal crises were to be avoided by the threat of "no bailouts," as enshrined in Article 125 of the Treaty on the Functioning of the European Union.

The Greek crisis exposed the time inconsistency of the "no bailout" promise. Faced with the threat of an imminent Greek sovereign default and high uncertainties about the direct and indirect costs of a default to the monetary union, eurozone member states found a quick fix to circumvent the "no bailout" clause: They granted a credit line of up to €80 billion in bilateral loans through a special vehicle, the Greek Loan Facility (GLF).[7] This exceptional vehicle was replaced first with the creation of a multilateral structure (the European Financial Stability Facility, EFSF) and then through a treaty establishing the permanent ESM.

Introducing a permanent facility for emergency financing amounted to a significant reform of the eurozone architecture. It added a supranational fiscal buffer for large crises and established a new regime of conditional bailouts. As noted above, three criteria for access to ESM funding are the main governors of this new regime, and should ensure that loans are extended only in cases of sustainable debt dynamics. However, the first and the third criteria are bound to be fulfilled in any crisis. If any default or restructuring is considered to raise doubt about the "integrity of the euro area as a whole," this test becomes meaningless as a commitment device.[8] Thus, the only real test is the analysis of debt sustainability.

The Greek debt crisis further affected the European financial architecture as loans from European partner countries were being restructured.

7. The volume eventually borrowed was €52.9 billion.

8. This interpretation was confirmed by the assessment of the Greek loan application of July 2015. The EU concluded that although direct financial risks of a Greek default were small, they would create "significant doubts on the integrity of the euro area as a whole, currently and in the future" (European Commission 2015, p. 5).

As presented in detail below, Europeans made a series of concessions and restructured their original loans multiple times. This official restructuring was silent but had a permanent impact on the Greek debt profile and on the institutions for crisis lending in Europe. The conditions of Greek loans were passed on to other crisis countries and became a de facto new ESM lending policy. As a consequence, European crisis lending conditions are now highly concessionary, with average maturities of up to 32.5 years, decade-long grace periods, and a pass-through of ESM funding costs to program countries. These lending terms are closer to those of the World Bank for long-term lending to low-income countries than to IMF-type, short-term balance-of-payments assistance.[9]

For the IMF, the first Greek program also brought about an important change in lending policies. The May 2010 stand-by program granted Greece exceptional access to draw €30 billion, more than 3,000 percent of its quota. The yardstick for granting such a high level of access was the debt sustainability criterion. Under the baseline scenario, the IMF projected Greece's public debt as a share of GDP to peak in 2013 at 149 percent and to gradually decline by 2020 to 120 percent, although it flagged many risks to this baseline scenario (IMF 2010a). On balance, the IMF considered debt to be sustainable over the medium term; however, it noted that the significant uncertainties "make it difficult to state categorically that this is the case with a high probability" (IMF 2010a, p. 20). Under the then-existing "exceptional access" policy, this statement would have precluded the IMF from approving the program without first requiring debt restructuring. The quick-fix solution was to introduce a "systemic exemption" from the rule due to the high risk of international spillovers.

This solution implies that the IMF could lend to insolvent countries, provided that spillovers are seen to be large.[10] The systemic exemption eventually became a permanent feature of IMF exceptional access policies

9. The maximum term on World Bank loans, under the International Bank for Reconstruction and Development, is 35 years, with an average weighted maturity of 20 years (World Bank 2014). Nonconcessional IMF loans are due much earlier, with final maturities from 3.25 to 5 years for Stand-By Arrangements and 4.5 to 10 years for the Extended Fund Facility; even for concessional loans, the IMF expects repayment within 8 to 10 years (IMF 2015c).

10. The fact that this constituted a change in policies was not obvious to the board and led to a heated discussion once one director pointed it out. The directors first thought that this exception would only be applied to Greece, but the Legal Department of the IMF explained that it would carry over to all member countries due to equal treatment requirements (IMF 2010b).

(IMF 2014c).[11] A former director of the IMF, Susan Schadler, put it like this: "The framework constraining the discretion of the IMF in severe debt crises broke down in its first serious test" (Schadler 2013, p. 14).[12] IMF staff members have proposed eliminating the systemic exemption on the grounds that it is inequitable and excessively open ended.[13] Instead of keeping the vague option of extending loans on grounds of systemic risk concerns, a recent staff proposal suggested that a one-time debt reprofiling (prolongation of maturities without reduction in principal or interest) should always be required in cases of doubtful debt dynamics (IMF 2014a).[14] By January 2016, the main shareholders of the IMF had accepted this argument, and the IMF announced a new lending policy abolishing the systemic exemption (IMF 2016).

However, the verdict on debt sustainability still constitutes an important condition for access to IMF and ESM lending. We next turn to the mechanics of these analyses.

I.C. The Mechanics of IMF and ESM Debt Sustainability Analysis

Both the IMF's and ESM's methodologies for analyzing debt sustainability require an analysis of the debt stock in a static framework using observed data about the current situation and in a dynamic framework using forecast data (IMF 2013b, 2013d; European Commission 2014). Forecasting requires a comprehensive macroeconomic model that at a minimum includes growth, inflation, interest, and exchange rates, as well as fiscal policies, and is therefore subject to uncertainty. Besides the benchmark assumptions, the data are also exposed to a series of robustness checks and stochastic analyses in which alternative data trajectories are considered.[15]

The results of these exercises, along with the static indicators, are then compared with a set of thresholds that designate an increased risk of debt

11. The systemic exemption has already been invoked 34 times for eurozone programs and reviews.

12. Schadler (2013) also discusses whether Greece really presented circumstances that warranted changing the lending criteria. She distinguishes between the immediate moment—when the counterfactual to the IMF's involvement would have entailed large systemic cost—and the continuing involvement of the IMF over the following years.

13. Moreover, it may increase the risk to the IMF's own resources and its seniority status, as evidenced by the default of Greece on payments to the IMF in the summer of 2015.

14. During the board discussion of this proposal, some directors preferred to keep the systemic exemption as a "pragmatic way to safeguard financial stability in an increasingly integrated world and avoid the perception of lack of evenhandedness" (IMF 2014b).

15. In particular, this includes negative shocks to the primary balance, real GDP growth, nominal interest rates, the exchange rate, and contingent liabilities.

distress. These thresholds are derived by running early warning systems, in the spirit of the "signaling approach" suggested by Graciela Kaminsky, Saul Lizondo, and Carmen Reinhart (1998); Kaminsky and Reinhart (1999); and Emanuele Baldacci and others (2011). A related alternative is the regression-based approach suggested by Aart Kraay and Vikram Nehru (2004). In the signaling approach, a signal of an impending crisis is triggered if the realized value of a set of macroeconomic and financial variables exceeds a critical value of the variable's distribution. If a signal is triggered and a crisis erupts in the following predefined projection period (such as the 24-month period in Kaminsky and Reinhart [1999]), the signal is recorded as a "good" positive; if no crisis occurs, it is counted as a "false" positive. Likewise, if there is a crisis but no signal has been recorded in the preceding projection period, every observation without a signal is recorded as a "false" negative. If no crisis erupts, and no signal was triggered, the observation counts as a "good" negative signal. The critical value of the distribution is chosen so as to minimize the equally weighted sum of false positive and false negative signals.

For instance, in the analysis by Kaminsky and Reinhart (1999), the threshold of the deficit-to-GDP ratio that minimizes the sum of false positives and false negatives is the 86th percentile of the historical deficit distribution. If a country's realized deficit in any given year exceeds the 86th percentile of that country's distribution of deficits, a crisis signal is triggered. While the original contributions by Kaminsky, Lizondo, and Reinhart (1998) and Kaminsky and Reinhart (1999) suggest country-specific distributions, the approach used today by the IMF (Baldacci and others 2011; IMF 2013b) and the ESM (Berti, Salto, and Lequien 2012; European Commission 2014) chooses thresholds based on the pooled distributions of all countries in the sample.

Similarly, the IMF thresholds for low-income countries are derived from a regression model in which a crisis indicator is regressed on the threshold variables; the maximum thresholds are then set so that the predicted crisis probability remains below predefined values (Kraay and Nehru 2004).

Table 1 shows the thresholds for the various frameworks. While the ESM does not distinguish between different countries, the IMF framework has different variables for market-access and low-income countries. In addition, the values for advanced economies and emerging markets are different, and within low-income countries values are further differentiated according to the institutional quality.

A country is only considered at low risk of debt distress if its debt stock and predicted future development do not exceed these thresholds,

Table 1. Thresholds for Risk of Debt Distress under Different Debt Sustainability Analysis Templates

	European Commission	*International Monetary Fund*								
		Market-access countries[a]						*Low-income countries*[b]		
	EU member states	*Advanced economies*			*Emerging markets*			*Weak*	*Medium*	*Strong*
Nominal debt[c]	90	85			70			38	56	74
Gross financing needs[d]	15	20			15				—	
Debt profile		*Low*	*Moderate*	*High*	*Low*	*Moderate*	*High*			
Bond spreads (basis points)[e]	<231[f] <276.6[g]	<400	400–600	>600	<200	200–600	>600		—	
External financing requirement[h]	—	<17	17–25	>25	<5	5–15	>15		—	
Foreign currency debt[i]	29.82		—		<20	20–60	>60		—	
Nonresident-held debt[j]	49.02	<30	30–45	>45	<15	15–45	>45		—	
Change in short-term debt[k]	2.76	<1	1–1.5	>1.5	<0.5	0.5–1	>1		—	
PPG external debt ratios[l]										
GDP	—		—			—		30	40	50
Exports	—		—			—		100	150	200
Revenues	—		—			—		200	250	300
PPG external debt service ratios[m]										
Exports	—		—			—		15	20	25
Revenues	—		—			—		18	20	22

Sources: European Commission (2014), International Monetary Fund (2013b, 2013d).

a. Market-access countries are designated as "advanced economies" or "emerging markets," based on their World Economic Outlook classification. Debt profile indicators further disaggregate market-access countries based on their levels of risk.

b. Countries qualify as "low-income" if they are eligible for the IMF's Poverty Reduction and Growth Trust (PRGT). A country is eligible for the PRGT if its annual per capita gross national income is below $1,215 and it does not have the capacity to access international financial markets on a durable and substantial basis. Low-income countries are further disaggregated by the quality of a country's policies and institutions, as measured by the World Bank's Country Policy and Institutional Assessment (CPIA). Countries with a stronger CPIA rating are considered to be more resilient to indebtedness, and therefore higher thresholds are applied.

c. Nominal debt level as a percent of GDP. For EU member states and market-access countries, nominal debt is expressed at face value; for low-income countries, at present value.

d. Sum of debt payments and individual payments or revenues (such as privatizations) less the primary balance in any given year, as a percent of GDP.

e. Government bond yield spreads relative to German and U.S. bonds of similar maturity.

f. 10-year benchmark.

g. 2-year benchmark.

h. Current account balance plus amortization of total short-term external debt at remaining maturity, as a percent of GDP.

i. Public debt held in foreign currency as a percent of total public debt.

j. Public debt held by nonresidents as a percent of total public debt.

k. Annual change in short-term public debt (at original maturity) as a percent of total public debt.

l. Present value of public and publicly guaranteed external debt, as a percent of various economic indicators.

m. Public and publicly guaranteed external debt service, as a percent of various economic indicators.

both under the benchmark scenario and under the robustness scenario with more negative assumptions. If the indicators exceed the thresholds in the baseline scenarios, the probability of debt distress is considered high. The middle ground is more ambiguous; a moderate risk rating is assigned if the thresholds are breached in the robustness scenarios but remain below the critical values under the baseline assumptions.

The general framework is therefore similar for all countries, in both the IMF framework and the ESM framework. However, the IMF analysis significantly differentiates between low-income countries and market-access countries along at least two dimensions that are not contained in the ESM analysis, as follows.

First, as the name suggests, market-access countries are assumed to borrow predominantly at market terms from market sources. This requires specific assumptions about the type and cost of market financing, including modeling the coupon, maturity, and currency structure of the debt. For low-income countries, borrowing from capital markets is considered an option, but many low-income countries rely mostly on official financing.

Second, for market-access countries, the stock of debt is considered at nominal values. The liabilities therefore only consist of the principal repayments, without taking into account coupon payments or the life of a debt instrument. The face-value measure of the debt stock, *FV*, is thus given by

$$FV = \sum_{t=0}^{T} A_t, \tag{1}$$

where A_t represents principal repayments in year *t*.

This is different for low-income countries, which receive most of their financing from official sources and whose debts are computed and analyzed in discounted present values. The present value, *PV,* is computed to include all discounted cash flows of the principal, A_t, and coupons, C_t:

$$PV = \sum_{t=0}^{T} \frac{C_t}{(1+d)^t} + \frac{A_t}{(1+d)^t}. \tag{2}$$

When computing meaningful present values, the key decision is choosing an appropriate discount rate, *d.* In the current IMF framework, *d* is set to a constant rate of 5 percent (IMF 2013c). This choice is justified by the fact that a more elaborate discounting model would increase the degrees of freedom in the analysis, thereby making cross-country comparisons more difficult. Conceivable alternatives that have been discussed in the

literature range from the London Interbank Offered Rate (commonly known as LIBOR) (Easterly 2001), to higher (constant) rates of between 7 and 10 percent (Chauvin and Kraay 2005; Andritzky 2006; Dikhanov 2006), to discounting based on the country's (or a reference country's) sovereign yield curve (Cruces and Trebesch 2013). In this paper, we do not take a stance on which of these approaches should be preferred; instead, to maximize comparability, we apply the IMF's discount rate of 5 percent.[16]

Both *FV* and *PV* only measure debt stocks, without taking into account funding pressures in any given year. To complement the analysis for this dimension, the gross financing needs, *GFN,* measure the difference between the debt service obligations and the government's available income for debt payments, the primary balance, *PB*:

$$GFN_t = A_t^* + C_t^* + A_t^N + I_t^N FV_{t-1}^N - PB_t,$$

and

$$FV_t^N = FV_{t-1}^N + GFN_t - A_t^N = \sum_{i=0}^{t} GFN_i - A_t^N,$$

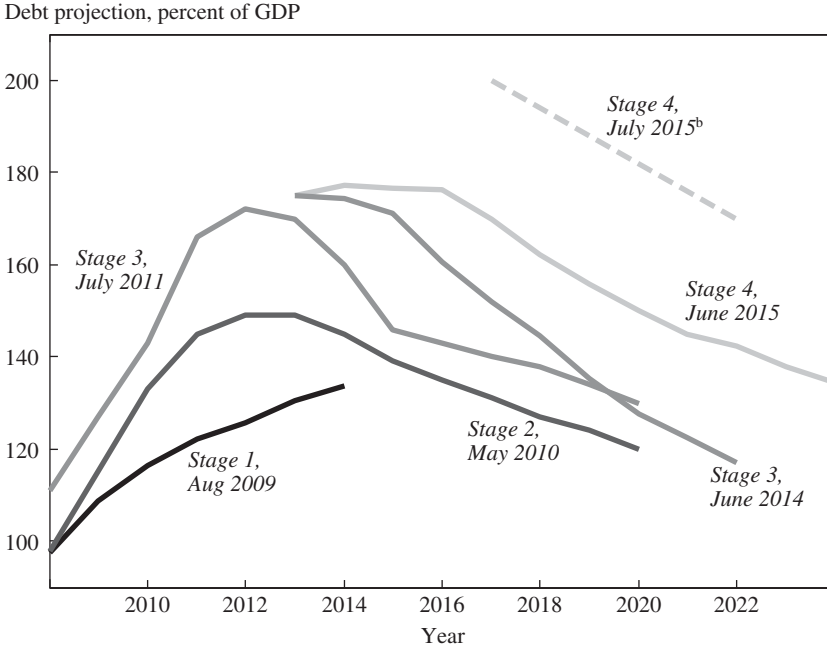
where A_t^* and C_t^* denote the debt repayments and interest that are already scheduled as of the date of the analysis, A_t^N denotes the repayment of newly issued debt to cover previous periods' *GFN,* I_t^N represents the interest burden on newly issued debt, and FV_t^N is the stock of newly issued debt. The future refinancing terms of the projected funding shortfalls—that is, the interest rate and maturity of newly issued debt—must be assumed. For instance, under an assumed maturity of 5 years, A_t^N will be equal to GFN_{t-5}.

II. Debt Sustainability Analysis in Greece: Past and Present Shortfalls

Ideally, the mechanical application of these tools should provide a clear result if a government's public finances are sustainable or not—independently of whether such a result is taken seriously in official lending frameworks. Yet in practice, significant uncertainties open a wide range of possible outcomes of the analysis, not least because it requires assumptions about the feasibility of future budgetary assumptions (House and Tesar 2015).

16. In the online appendix, we report alternative results based on Greek and German market-discount rates. Online appendixes to all papers in this volume may be found on the *Brookings Papers* webpage, www.brookings.edu/bpea, under "Past Editions."

Figure 1. Four Stages of (Mis-)Diagnosed Greek Debt Sustainability[a]



Source: International Monetary Fund (2009, 2010a, 2011, 2015a, 2015b).

a. Shows the debt projections as a percent of GDP from the IMF's debt sustainability analyses at the four stages described in the text.

b. The latest publication from July 2015 did not contain a full analysis, but only two data points: a peak at close to 200 percent within the next two years, and a reduction to 170 percent in 2022. The dashed line represents the lack of full data.

II.A. Serial Misdiagnoses and Restructuring

Indeed, over the course of the Greek crisis, debt dynamics have been repeatedly underestimated. We divide the recent history into four stages, with the first stage starting in the fall of 2009 (figure 1). At that time, Greece was already in an "Excess Deficit Procedure" for breaching the limits of the Maastricht Treaty, and it had committed to bringing its deficit back to 3 percent of GDP during the coming year. In the summer of 2009, the IMF estimated the current deficit at 6.2 percent and warned that debt dynamics would become unsustainable unless policies were radically changed (IMF 2009). The debt-to-GDP ratio was expected to rise above 100 percent in 2009 and to increase further, to more than 120 percent, within 2 years. The IMF analysis concluded that fiscal consolidation was immediately required to achieve sustainability but that fairly modest adjustments of 1.5 percent

of GDP would be sufficient. In October 2009, Greece revealed new deficit estimates of up to 12.5 percent, at the same time acknowledging that it had misreported previous numbers, which also turned out to be substantially higher.[17] Concerns about fiscal sustainability deepened and triggered a confidence crisis.

The second stage began in May 2010 with the first joint bailout by eurozone governments and the IMF. While the dynamic was seen as considerably more negative than before and the debt stock was seen to peak at close to 150 percent of GDP, the verdict was still that the debt was sustainable, if not with high probability. The official loans were justified by invoking the systemic exemption for the first time and trusting that debt restructuring would not be necessary.

The third stage was the time of reckoning, reached in mid-2011, when official sector creditors acknowledged that debt restructuring was unavoidable (IMF 2011). The well-publicized and well-documented part of this stage was the restructuring of private debt, which took place in March 2012. The process involved retroactively changing bond contracts by legislative action and a good measure of coercion by governments on financial institutions; but the result was a high participation rate and a severe haircut, with present value reduction of over 60 percent (Zettelmeyer, Trebesch, and Gulati 2013).

The official sector restructuring was more silent. Over time, European public sector loans were restructured, deeply and repeatedly. Table 2 shows the timeline for Greek debt restructurings through the two main public loan vehicles for Greece, the GLF and the EFSF. Interest rates on bilateral loans in the GLF were lowered in three steps between 2010 and 2013, reducing the interest margin over the floating 3-month Euro Interbank Offered Rate (commonly known as EURIBOR) from 300–400 basis points to 50 basis points. Even more pronounced were the extensions of the grace period, from 3 to 10 years, and of the maturity, from 5 to 30 years. EFSF loan conditions were restructured in a similar way, most importantly by almost

17. Moreover, Eurostat noted that it could not verify the new figures and flagged the risk of further upward revisions. The ministers of the Economic and Financial Affairs Council immediately mandated the European Commission to investigate. The resulting report concludes that the quality and the governance of Greek fiscal statistics are seen as insufficient, noting that "revisions of this magnitude . . . have been extremely rare in other EU Member States, but have taken place for Greece on several occasions" (European Commission 2010, p. 3).

18. The conditions of EFSF lending were amended accordingly, and the other EFSF program countries—Ireland and Portugal—similarly benefited from debt relief in the form of significantly increased maturities. The maturity on Ireland's loan was increased from 2016–29 to 2029–42, and on Portugal's loan from 2015–38 to 2025–40, increasing the average weighted maturity to more than 20 years (EFSF 2013a, 2013b).

Table 2. Greek Loan Conditions over Time, 2010–14

Greek Loan Facility[a]	*May 9, 2010*	*June 14, 2011*	*Feb 27, 2012*	*Dec 19, 2012*[b]
Margin:	300–400 basis points	200–300 basis points	150 basis points	50 basis points
Grace period:	3 years	4.5 years	10 years	10 years
Maturity:	5 years	10 years	15 years	30 years

European Financial Stability Facility[a]	*March 1, 2012*	*Dec 12, 2012*[b]
Margin:	0 basis points	0–200 basis points
Fees:	>15 basis points	>5 basis points
Grace period:	0 years	10 years
Average maturity:	17.5 years	32.5 years

Private	*March 9, 2012*	*Dec 12, 2012*
FV reduction:	53.5 percent	64.6 percent
PV reduction:	64.4 percent	61.4 percent

Sources: For the Greek Loan Facility, EU (2010, 2011, 2012a, 2012b); European Financial Stability Facility (2013a, 2013b); Eurogroup (2012); and Zettelmeyer, Trebesch, and Gulati (2013).

a. The first entries for the Greek Loan Facility and the European Financial Stability Facility denote the original terms of the loans. The remaining entries denote the terms of subsequent amendments.

b. The November 2012 agreement of the Eurogroup in which the Greek Loan Facility and European Financial Stability Facility restructurings of December 2012 were announced contained further measures to ease the Greek debt burden that were not part of the implemented agreements. In particular, these include a commitment to pass on profits from the bond purchases under the European Central Bank's securities markets program, and that further adjustments of the loan conditions would be considered conditional on the successful implementation of the reform program.

doubling the average maturity of the loans to more than 30 years.[18] ESM lending policies were later aligned.

After the combined private and official debt relief, the troika concluded that Greek debt was finally sustainable. Moreover, the assessment became gradually more optimistic, and the IMF released a new debt projection in June 2014. For the first time, no further increase in the debt stock was projected—it seemed that the peak had been left behind. By the beginning of 2015, the troika viewed Greece as being on a good path. In its request to the German parliament for an extension of the Greek program, the German Ministry of Finance justified the extension with a "confirmation of debt sustainability" by the European Commission, and explained that the "debt sustainability has improved since the last program review of April 2014" (German Ministry of Finance 2014, p. 4).

The fourth and ongoing stage is characterized by conflict between official creditors about debt sustainability, which at the time of this writing remains unresolved. The IMF has made debt relief a condition for participation in a third Greek program. In July 2015, it published two new debt projections within a short time. It then argued that the systemic exemption can no longer be invoked for Greece and that it will not participate in funding a new program unless there is further debt restructuring on the European side. The Europeans have decided to go ahead with financing without restructuring the existing loans again, leaving burden sharing within the official sector an unresolved conflict.

II.B. Uncertainties in the Analysis of Greece's Debt Stock

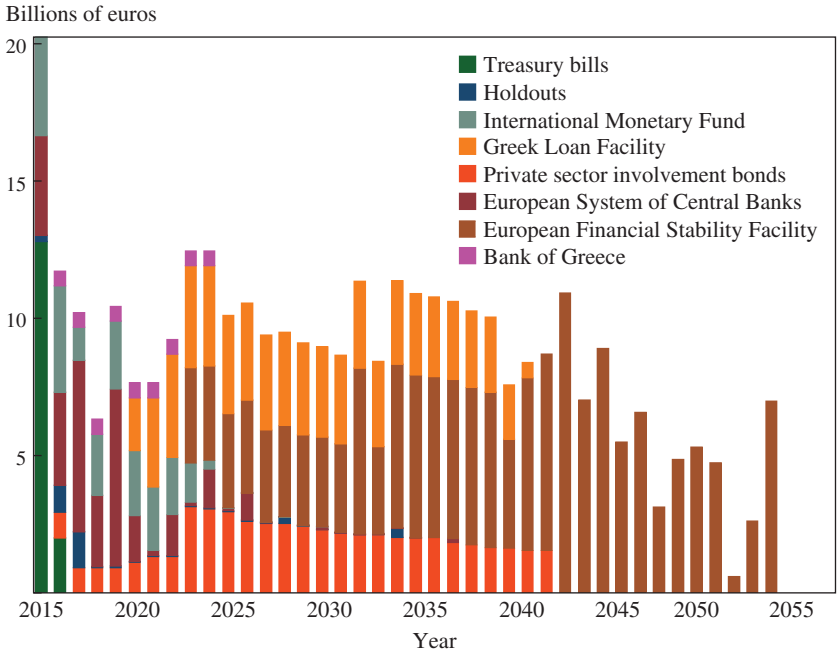
While Greece had been exclusively relying on private financing between the introduction of the euro and the start of the European sovereign debt crisis in 2009, Greek debt today is dominated by official loans. As of the end of July 2015, before the third program had been negotiated, more than 80 percent of its current outstanding debt was owed to official creditors. The average maturity was 15.7 years, with an average interest rate of 2.7 percent.[19]

Figure 2 shows the debt repayment profile of Greece by creditor.[20] Only the relatively large amount of short-term debt (Treasury bills) and the remaining holdout bonds that were not restructured in 2012 require repayments to private investors within the next 8 years. Afterward, the remaining

19. See Public Debt Management Agency (2015). To put this in perspective, the average maturity terms of Italian and French sovereign debt are 6.5 and 6.9 years, respectively (Italy, Department of the Treasury 2015; France, Treasury Agency 2015).

20. Not including the new ESM program.

Figure 2. Greek Debt Repayment Profile (Face Value) by Creditor, 2015–54



private sector involvement (PSI) bonds amortize over a period of 20 years, stretched out through the maturity extensions of the 2012 debt restructuring. The bulk of the debt is owed to the EFSF; to other eurozone governments through the GLF; to the IMF; to the ECB; and to other members of the European System of Central Banks. Notably, the official European loans through the GLF and the EFSF only start becoming due in 2020 and 2023, respectively, and repayments are stretched out until 2054.

As explained above, the official loans are extended at highly favorable terms (table 3). This generates a significant element of concessionality. Using the discount rate of 5 percent to compare the face value of the EFSF and GLF loans to their present value reveals considerable "grant elements" of up to 61 percent. On average, the Greek debt stock contains a grant element of 37 percent.[21]

The recently negotiated third program over €86 billion is likely to increase this concessionality. The new ESM program will have the same 32.5 year

21. A previous version of these and the following computations can be found in Schumacher and Weder di Mauro (2015).

Table 3. Greek Debt Composition

Debt	*Face value*[a] *(billions of euros)*	*Interest*[b] *(billions of euros)*	*Present value*[c] *(billions of euros)*	*Grant element (percent)*[d]
Treasury bills	14.8	—	14.6	n/a
Private sector involvement bonds	30.5	20.7	26.8	12
Holdout bonds	2.8	0.7	2.9	–2
Bonds held by the Eurosystem[e]	23.6	4.9	24.2	–2
International Monetary Fund	19.5	2.5	18.6	5
European Financial Stability Facility	131.0	27.8	51.4	61
Greek Loan Facility	52.9	13.5	33.2	37
Bank of Greece[f]	4.8	—	3.8	n/a
Total (without new program)	280.1	70.0	175.6	37
Percent of GDP	156		98	
New program[g]				
European Stability Mechanism	77.8	31.7	31.9	59
International Monetary Fund	8	1	8	8
Total new program	86.0	32.7	39.5	54.1
Total	366.1	102.6	215.1	41.2
Percent of GDP	204		120	

Sources: IMF (2013d), for the definition of the grant element; for the data, the various data sources cited in the paper (IMF, GLF, EFSF).

a. Sum of principal payments, as in equation 1.

b. Undiscounted sum of interest payments due to each creditor.

c. Discounted sum of principal and interest payments, as in equation 2.

d. Defined as (face value − present value) / face value.

e. Includes bonds held by the European Central Bank, the European Investment Bank, and various national central banks.

f. Assuming constant amortization.

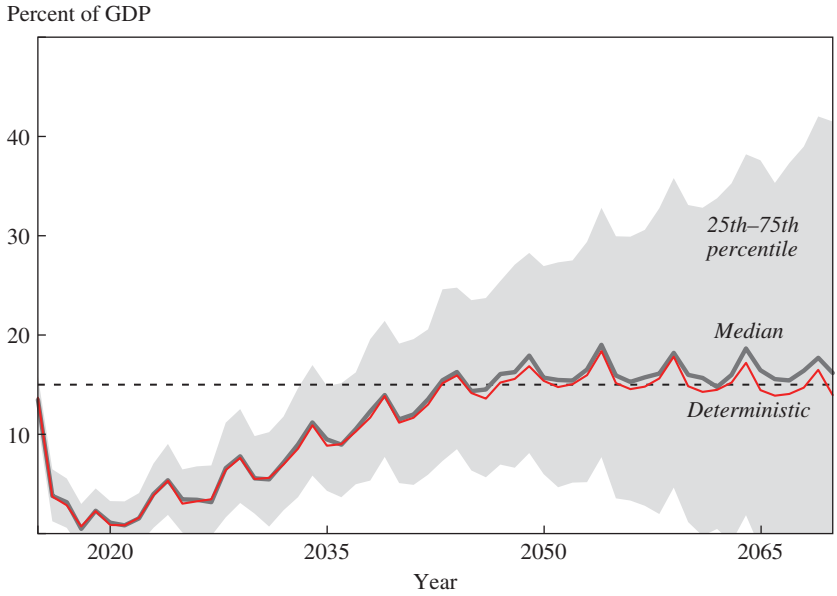
g. Assuming the new program is identical to the previous programs with respect to the share of IMF and European lending.

average maturity as the EFSF loans, with amortizations beginning in 2034, and similarly favorable interest rates. Assuming that the IMF will contribute circa 10 percent of the total volume, in line with the currently outstanding share of IMF and European commitments, and request a maturity of 5 years, the average grant element will rise to more than 40 percent.

II.C. Uncertainties in the Analysis of Flows

In its most recent debt sustainability analysis, the IMF has expressed a similar judgment. Furthermore, the IMF has acknowledged that the nominal gross debt-to-GDP ratio is no longer a meaningful metric to evaluate

Figure 3. Gross Financing Needs over the Long Term, 2015–70



sustainability (IMF 2015a, p. 11). The recent analysis suggests that gross financing need (GFN) should be looked at instead, as these needs capture the funding pressure in any given year. Using values similar to those indicated in the recent IMF debt sustainability analysis (annual real GDP growth of 1.75 percent, average interest rate of 6 percent, average maturity of 5 years, and primary balance of 3.5 percent), the GFN is projected to increase relative to GDP above critical thresholds of 15 to 20 percent from the mid-2030s onward (figure 3). In the short to medium terms, however, the GFN remains below this critical value.

Two factors explain the increase in GFN from the 2030s onward. First, the European loan repayment schedules fall together with redemptions of private sector bonds after the grace periods on the official loans have ended in the mid-2020s. However, even under modest growth assumptions, the annual obligations do not exceed €15 billion, so in isolation they remain well below 15 percent of GDP. The second, more crucial factor in this analysis is therefore the refinanced debt from the relatively low GFN in the coming 10 to 15 years. Every annual funding gap is refinanced at the assumed market interest rate of 6 percent with a maturity of 5 years. Over the very-long-term horizon considered in this analysis, the compounding

of these relatively expensive terms (as compared with today's low-interest environment) leads to the very high GFN displayed in figure 3.

For instance, the total projected debt payments in 2050 amount to €62.1 billion under the chosen parameters. Of these, €42.6 billion consists of the repayments of debt projected to be issued in 2045 to cover the financing gap in that year. Another €14.1 billion consists of the projected interest due on the outstanding debt stock, and only €5.3 billion are the interest and principal payments scheduled to the EFSF as of 2015. Together with a projected primary surplus of €11.5 billion, the 2050 GFN comes down to €50.6 billion. In other words, only about 10 percent of the GFN for that year is based on terms known today, and the remainder rests on the accuracy of the assumptions about market conditions over the coming 35 years.

This long projection horizon implies considerable uncertainty. The gray area in figure 3 displays the results of a simple Monte Carlo simulation of the projected GFN, showing the 25th and 75th percentile of realizations.[22] For 2050, the interquartile range reaches from 6 to 27 percent. The policy implications of these two outcomes for the requirement of debt restructuring today would of course be fundamentally different. While a low GFN would imply healthy expected finances, a value of 27 percent would even breach the higher bound of the GFN range mentioned in table 1 by a wide margin.

The IMF has made debt operation a precondition for continued involvement in Greece and has proposed that restructuring could take the form of doubling the maturities on the European loans. However, due to the significant grant element of the EFSF and GLF loans, a pure reprofiling of this part of Greek debt will achieve a relatively smaller reduction in the present-value debt stock than an extension of maturities of earlier liabilities. Average maturities on the GLF and EFSF loans are already more than 30 years. For every €1 due in 30 years, a doubling of maturities reduces the present value of this liability by only 19 cents. Conversely, extending the term of €1 coming due next year by only 10 years reduces the present value of that obligation by 38 cents.

22. The analysis rests on 1,000 draws of independent and identically distributed shocks to the growth rate (g), the interest rate (r), and the primary balance (pb), each drawn from a normal distribution with standard deviations matched to the historical data between 2001 (entry into the eurozone) and 2009 (last precrisis year). Specifically, for a normal distribution $N(\mu, \sigma^2)$, the shock distributions are $g \sim N(2.7, 3.5)$, $r \sim N(4.6, 0.6)$, and $pb \sim N(-2.3, 3.6)$, where μ and σ^2 are measured in percent.

A debt operation that only extended the European loans would therefore be relatively less efficient in achieving present-value debt stock reductions than would a restructuring that included shorter-term liabilities. Furthermore, it would not ease financing needs until the mid-2020s, when those loans start coming due.

An extension of GLF or EFSF maturities would indeed bring down the projected GFN by reducing the amount that has to be refinanced at the assumed unfavorable market terms. But the effectiveness of such an operation rests strongly on the accuracy of the macroeconomic, financial, and fiscal projections. If market interest rates remain elevated for Greece, GDP growth remains sluggish, or a primary balance of 3.5 percent proves elusive over the next four decades, any return to private sector funding will be difficult. Conversely, a more positive outcome would make a maturity extension of the official loans obsolete.

III. Policy Implications

The analysis of this paper has policy implications on three levels. First, there are implications for the ongoing debate about a restructuring of Greek debt. We show that the nominal debt stock projections paint far too bleak a picture of the actual burden. Evaluated in present-value terms, Greek debt stands at about 100 percent and will rise to about 120 percent under the new program, which is not exceptionally high for advanced countries. Nevertheless, even in present-value terms, Greece still breaches the thresholds of the standard debt sustainability analysis for both market-access countries and lower-income countries. The projection of GFN over the short to medium runs, however, does not provide significant reasons for debt relief. The long-term projections, while sending signals of critically high funding pressure, are marked by very large uncertainties and thus are not a reliable basis for deciding the restructuring need.

Consequently, the debt stock and projected payment flows show that, despite the extraordinary amount of private and public debt relief Greece has already received, further debt restructuring may be advisable, although that conclusion is far less certain than commonly argued.[23] If required, a debt operation should focus on the horizon over which the payment flow projections are relatively reliable and a restructuring would be relatively

23. See House and Tesar (2015) for an analysis of the required budgetary adjustments if no debt reduction is implemented.

more efficient in reducing the debt stock. But in the short to medium runs, the repayments are mostly owed to the IMF and the ECB, and only a further extension of grace periods by the European partners would reduce the risk of default on their loans. Thus, an efficient debt restructuring will have to answer the question of burden sharing and seniority within the official sector first.

Furthermore, a mere extension of maturities without a reduction in the nominal value of long-run obligations will only extend the interdependencies of Greece and its European creditors. In their paper in this volume, Reinhart and Trebesch (2015) show that such long-term financial dependencies create significant political tensions. Political decisionmakers in the current debate should be well aware of such frictions when engaging in debt operations, which can prolong such potentially bruising negotiations for decades to come. This political economy argument speaks for outright debt relief rather than further prolongations. But a second historical lesson from Reinhart and Christoph Trebesch (2015) is that Greece has been prone to quickly overborrow again as soon as the previous debt crisis has been overcome. This suggests that debt relief should only be granted after Greece has demonstrated that it is able and willing to break away from the historical pattern, and lends support for a process like that for the Highly Indebted Poor Countries, whereby multilateral debt forgiveness is granted after an extended track record of good policy has been established.

The second set of implications concerns the mechanics of assessing debt sustainability in the official sector. Both the ESM and the IMF still apply the market-access framework to Greece, not taking into account the effective present-value debt relief that has already been granted. This is especially paradoxical in the case of the ESM, since the grant element of up to 60 percent is only contained in European loans. Nevertheless, the headline number of roughly 200 percent of nominal debt-to-GDP stock is still used in public and in negotiations. While the low-income country framework of the IMF and World Bank does account for grant elements in official lending, it does not seem appropriate for a case like Greece. In particular, the low-income country framework sets different levels for debt sustainability depending on the quality of institutions and policies. This may be sound in principle, but not feasible in practice inside the eurozone.

Finally, there are broader implications for the European monetary union as well as the international monetary system. The Greek debt crisis has profoundly changed the architecture of the eurozone. In addition to spurring reforms in EU fiscal governance, it has led to the creation of the ESM as a permanent crisis-lending mechanism. Moreover, Greece contributed to

softening the ESM's lending framework and to transforming it from an institution like the IMF to one like the World Bank. The importance of this last step has not been sufficiently recognized. On one hand, it has implications for the ongoing debate on fiscal union, since the large grant element in European crisis lending has added a fiscal buffer. On the other hand, the repeated softening of lending conditions signals that the ESM has commitment problems and strengthens the case for establishing an effective regime for sovereign debt restructuring in the eurozone.

In contrast, the IMF's recent reforms seek to reverse the softening of its crisis lending framework. By abandoning the systemic exemption introduced in 2010, future lending decisions by the IMF should take the results of debt sustainability analyses more seriously again. Linking emergency loans to an obligatory maturity extension of existing debt if it cannot be assessed to be sustainable with high probability may help to overcome commitment problems. However, negative externalities can always be expected in cases serious enough to require exceptional access to the IMF. The framework will therefore have to stand the test of future crises to reveal the true robustness of its rules.

ACKNOWLEDGMENTS We thank the editors, the seminar participants at the 2015 *Brookings Papers* Fall Meeting, and also Reza Baqir, Antonio Fatas, Nicolas Sauter, Susan Schadler, Christoph Trebesch, and Jeromin Zettelmeyer for very helpful comments.

References

Aguiar, Mark, and Manuel Amador. 2014. "Sovereign Debt." In *Handbook of International Economics* 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Amsterdam: North-Holland.

Andritzky, Jochen. 2006. *Sovereign Default Risk Valuation: Implications of Debt Crises and Bond Restructurings.* Berlin: Springer.

Baldacci, Emanuele, Iva Petrova, Nazim Belhocine, Gabriela Dobrescu, and Samah Mazraani. 2011. "Assessing Fiscal Stress." Working Paper no. 11/100. Washington: International Monetary Fund.

Berti, Katia, Matteo Salto, and Matthieu Lequien. 2012. "An Early-Detection Index of Fiscal Stress for EU Countries." Economic Paper no. 475. Brussels: European Commission.

Buchheit, Lee C., Anna Gelpern, Mitu Gulati, Ugo Panizza, Beatrice Weder di Mauro, and Jeromin Zettelmeyer. 2013. "Revisiting Sovereign Bankruptcy." Washington: Brookings Institution, Committee on International Economic Policy and Reform.

Chauvin, Nicolas Depetris, and Aart Kraay. 2005. "What Has 100 Billion Dollars Worth of Debt Relief Done for Low Income Countries?" Working Paper. http://siteresources.worldbank.org/DEC/Resources/ChauvinKraayWhatHasDebtReliefAccomplishedSept2005.pdf

Cruces, Juan J., and Christoph Trebesch. 2013. "Sovereign Defaults: The Price of Haircuts." *American Economic Journal: Macroeconomics* 5, no. 3: 85–117.

Dias, Daniel A., Christine Richmond, and Mark L. J. Wright. 2014. "The Stock of External Sovereign Debt: Can We Take the Data at 'Face Value'?" *Journal of International Economics* 94, no. 1: 1–17.

Dikhanov, Yuri. 2006. "Historical PV of Debt in Developing Countries, 1980–2002." Report no. 36153. Washington: World Bank.

Easterly, William. 2001. "Growth Implosions, Debt Explosions, and My Aunt Marilyn: Do Growth Slowdowns Cause Public Debt Crises?" Policy Research Working Paper no. 2531. Washington: World Bank.

Eaton, Jonathan, and Raquel Fernandez. 1995. "Sovereign Debt." In *Handbook of International Economics* 3, edited by Gene M. Grossman and Kenneth Rogoff. Amsterdam: North-Holland.

Eaton, Jonathan, and Mark Gersovitz. 1981. "Debt with Potential Repudiation: Theoretical and Empirical Analysis." *Review of Economic Studies* 48, no. 2: 289–309.

EFSF (European Financial Stability Facility). 2013a. "Amendment Agreement Relating to the Loan Facility Agreement, Financial Assistance Facility Agreement, and Master Financial Assistance Facility Agreement between European Financial Stability Facility, the Portuguese Republic as Beneficiary Member State, and Banco de Portugal." Lisbon and Luxembourg.

———. 2013b. "Amendment Agreement Relating to the Loan Facility Agreement, First Financial Assistance Facility Agreement, Second Financial Assistance Facility Agreement, and Master Financial Assistance Facility Agreement

between European Financial Stability Facility, Ireland as Beneficiary Member State, and Central Bank of Ireland." Dublin and Luxembourg.

EU (European Union). 2010. "Loan Facility Agreement between the Member States Whose Currency Is the Euro and the Hellenic Republic." Brussels.

———. 2011. "Amendment to the Loan Facility Agreement of June 2011." Brussels.

———. 2012a. "Amendment to the Loan Facility Agreement of February 2012." Brussels.

———. 2012b. "Amendment to the Loan Facility Agreement of December 2012." Brussels.

Eurogroup. 2012. "Eurogroup Statement on Greece." Brussels. http://www.consilium.europa.eu/press/press-releases/2012/11/pdf/Eurogroup-statement-on-Greece-27-11/

European Commission. 2010. "Report on Greek Government Deficit and Debt Statistics." Brussels.

———. 2014. "Assessing Public Debt Sustainability in EU Member States: A Guide." Occasional Paper no. 200. Brussels.

———. 2015. "Greece: Request for Stability Support in the Form of an ESM Loan." Brussels.

France, Treasury Agency. 2015. *Monthly Bulletin,* no. 302. Paris.

German Ministry of Finance. 2014. "Request of the Federal Ministry of Finance: Grants in Favor of Greece; Technical Extension and Continuation of the Stability Support" [in German]. Berlin: German Bundestag.

House, Christopher, and Linda Tesar. 2015. "Greek Budget Realities: No Easy Option." In present volume of *Brookings Papers on Economic Activity.*

IMF (International Monetary Fund). 1995. "IMF Approves US$17.8 Billion Stand-By Credit for Mexico." Press Release no. 95/10. Washington.

———. 2002. "Access Policy in Capital Account Crises." Washington.

———. 2004. "Review of Exceptional Access Policy." Washington.

———. 2008. "Reform of Quota and Voice in the International Monetary Fund: Report of the Executive Board to the Board of Governors." Washington.

———. 2009. "Greece: 2009 Article IV Consultation—Staff Report; Staff Supplement; Public Information Notice on the Executive Board Discussion; and Statement by the Executive Director for Greece." Country Report no. 09/244. Washington.

———. 2010a. "Greece: Staff Report on Request for Stand-By Arrangement." Country Report no. 10/110. Washington.

———. 2010b. "Minutes of Executive Board Meeting 10/45-1." Washington.

———. 2011. "Greece: Fourth Review under the Stand-By Arrangement and Request for Modification and Waiver of Applicability of Performance Criteria." Country Report no. 11/175. Washington.

———. 2013a. "Sovereign Debt Restructuring: Recent Developments and Implications for the Fund's Legal and Policy Framework." Washington.

———. 2013b. "Staff Guidance Note for Public Debt Sustainability Analysis in Market-Access Countries." Washington.

———. 2013c. "Unification of Discount Rates Used in External Debt Analysis for Low-Income Countries." Washington.

———. 2013d. "Staff Guidance Note on the Application of the Joint Bank-Fund Debt Sustainability Framework for Low-Income Countries." Washington.

———. 2014a. "The Fund's Lending Framework and Sovereign Debt: Preliminary Considerations." Washington.

———. 2014b. "IMF Executive Board Discusses the Fund's Lending Framework and Sovereign Debt." Press Release no. 14/294. Washington.

———. 2014c. *Selected Decisions and Selected Documents of the International Monetary Fund,* no. 37. Washington.

———. 2015a. "Greece: Debt Sustainability Analysis (Preliminary Draft)." Country Report no. 15/165. Washington.

———. 2015b. "Greece: An Update of IMF Staff's Preliminary Public Debt Sustainability Analysis." Country Report no. 15/186. Washington.

———. 2015c. "IMF Lending." Factsheet. Washington. https://www.imf.org/external/np/exr/facts/howlend.htm

———. 2016. "IMF Executive Board Approves Exceptional Access Lending Framework Reforms." Press Release no. 16/31. Washington.

Italy, Department of the Treasury. 2015. "Government Debt: Breakdown by Instrument as of July 31, 2015." Rome: Ministry of Economy and Finances.

Kaminsky, Graciela, Saul Lizondo, and Carmen M. Reinhart. 1998. "Leading Indicators of Currency Crises." *IMF Staff Papers* 45, no. 1: 1–48.

Kaminsky, Graciela L., and Carmen M. Reinhart. 1999. "The Twin Crises: The Causes of Banking and Balance-of-Payments Problems." *American Economic Review* 89, no. 3: 473–500.

Kazarian, Paul. 2015. "Greece Sovereign Debt: Asking the Right Questions." Speech given at Conference on Rethinking Sovereign Debt. Munich: CESifo.

Kraay, Aart, and Vikram Nehru. 2004. "When Is Debt Sustainable?" Policy Research Working Paper no. 3200. Washington: World Bank.

Levy Yeyati, Eduardo, and Ugo Panizza. 2011. "The Elusive Costs of Sovereign Default." *Journal of Development Economics* 94, no. 1: 95–105.

Panizza, Ugo, Federico Sturzenegger, and Jeromin Zettelmeyer. 2009. "The Economics and Law of Sovereign Debt and Default." *Journal of Economic Literature* 47, no. 3: 651–98.

Public Debt Management Agency. 2015. *Hellenic Republic Public Debt Bulletin*, no. 77. Athens.

Reinhart, Carmen M., and Christoph Trebesch. 2015. "The Pitfalls of External Dependence: Greece, 1829–2015." In present volume of *Brookings Papers on Economic Activity.*

Sandleris, Guido. 2015. "The Costs of Sovereign Defaults: Theory and Empirical Evidence." Working Paper no. 02/2015. Buenos Aires: Torcuato di Tella University Business School, Finance Research Centre (CIF).

Schadler, Susan. 2013. "Unsustainable Debt and the Political Economy of Lending: Constraining the IMF's Role in Sovereign Debt Crises." CIGI Paper no. 19. Waterloo, Canada: Centre for International Governance Innovation.

Schumacher, Julian, and Beatrice Weder di Mauro. 2015. "Debt Sustainability Puzzles: Implications for Greece." London: Centre for Economic Policy Research, VoxEU. http://www.voxeu.org/article/debt-sustainability-puzzles-implications-greece

Serafeim, George. 2015. "Greece's Debt: Sustainable?" Harvard Business School, Case no. 115-063. Cambridge, Mass.: Harvard Business Publishing.

Tomz, Michael, and Mark L. J. Wright. 2013. "Empirical Research on Sovereign Debt and Default." *Annual Review of Economics* 5: 247–72.

Van Rijckeghem, Caroline, and Beatrice Weder. 2009. "Political Institutions and Debt Crises." *Public Choice* 138, nos. 3–4: 378–408.

World Bank. 2014. "IBRD/IDA Management's Discussion & Analysis and Financial Statements." Annual Report. Washington.

Zettelmeyer, Jeromin, Christoph Trebesch, and Mitu Gulati. 2013. "The Greek Debt Restructuring: An Autopsy." *Economic Policy* 28, no. 3: 513–63.

CARMEN M. REINHART
Harvard University

CHRISTOPH TREBESCH
University of Munich

The Pitfalls of External Dependence: Greece, 1829–2015

ABSTRACT Two centuries of Greek debt crises highlight the pitfalls of relying on external financing. Since its independence in 1829, the Greek government has defaulted four times on its external creditors—with striking historical parallels. Each crisis is preceded by a period of heavy borrowing from foreign private creditors. As repayment difficulties arise, foreign governments step in, help to repay the private creditors, and demand budget cuts and adjustment programs as a condition for the official bailout loans; political interference from abroad mounts, and a prolonged episode of debt overhang and financial autarky follows. We conclude that these cycles of external debt and dependence are a perennial theme of Greek history, as well as in other countries that have been "addicted" to foreign savings. At present, there is considerable evidence to suggest that a substantial haircut on external debt is needed to restore the economic viability of Greece. Even with that, a policy priority for the country should be to reorient, to the extent possible, toward domestic sources of funding.

The history of Greece is a narrative of debt, default, and external dependence. In 1952, the Greek-Canadian historian L. S. Stavrianos noted that since their independence, "the Greek people have had to bear a crushing foreign debt that has literally sucked their lifeblood" (Stavrianos 1952, p. 25). This graphic statement could well have been written 60 years later, in 2012, when Greece was in the midst of its fourth sovereign debt crisis. Or it could have been written 60 years earlier, on the eve of the second sovereign default. This paper documents the recurring patterns of sovereign default in Greece with the aim of gaining insights into possible solutions to the current crisis.

Our main conclusion is that the composition of Greek sovereign debt (external versus internal), and not just its levels, played a central role in explaining the country's historical default episodes, as well as its current predicament. Over the past 200 years, the tilt toward foreign borrowing in Greece (by the public and private sector) has resulted in repeated crises and sudden reversals (stops) of capital flows. We highlight that the consequences of the boom-bust cycles in external borrowing were not only economic, but political as well. The defaults resulted in prolonged bouts of heavy political interference from abroad, mainly aimed at assuring the repayment of bailout loans. The events since 2010 are neither new nor unique in Greek history.

There are relatively few papers on the unfolding Greek crisis that take a longer historical perspective. In this paper, we focus on Greece in the long-run, though our data and archival work is part of a much broader research agenda on the history of sovereign lending, default, and haircuts, which covers all debtor countries over the past 200 years (see Meyer, Reinhart, and Trebesch [ongoing work]).

The evidence we present reveals striking historical parallels between the past and the present. Most surprising are the close similarities in the crisis resolution process. For example, we find that Greece has been bailed out many times before, coupled with heavy conditionality and externally imposed adjustment programs. We also find that earlier Greek defaults have been similarly protracted, and that much of the bailout money was used to service old privately held debt. In each crisis, the country's external creditors (both official and private) initially refused to accept haircuts, but agreed to them eventually, sometimes after decades of fruitless negotiations and failed interim agreements. These insights speak to the current debate on how to address Greece's current debt overhang.

More generally, the role of external versus domestic borrowing remains comparatively understudied in connection to economic crises. Carmen Reinhart and Kenneth Rogoff (2009) take up this theme when discussing the literature at large. In the case of Greece, the debate has focused on other issues, such as debt sustainability, contagion effects, the need for reform, and the associated political economy problems. The fact that the ongoing crisis is very much an *external* debt crisis has been largely overlooked.[1]

1. We concur with Gros (2013) and Sinn (2014), that the crisis in periphery Europe is not so much a crisis of public debt, but rather a crisis of external debt, involving all the problems that come with an external crisis (in particular sudden stops, balance sheet effects, and cross-border disputes between creditors and debtors). In this regard, the analysis by Eichengreen and others (2014), which compares the eurozone crisis to Latin America's lost decade in the 1980s, is exactly on point.

The financial history of Greece also serves as a broader precautionary note for other countries that are "addicted" to foreign savings. Periods with external dependence and financial openness were often periods of volatility and crises, such as Latin America from the 19th century, but also in places like China, Portugal, or Spain, until these turned inward in the second half of the 20th century. Beyond much of Latin America, large emerging markets, such as Indonesia, Turkey, and parts of eastern Europe rely heavily on foreign saving.[2] We realize that our message that external debt implies important risks stands in contrast to recent calls to unravel the "deadly embrace" between governments and domestic banks, mainly by reducing the home bias in sovereign debt holdings (Corsetti and others 2015). Yet bank portfolios were almost entirely domestic from 1945 to 1980, the period in history with the fewest banking and debt crises (Reinhart and Rogoff 2009). Also, the most prosperous and financially stable period in Greek history, between the 1950s and 2000, was a period with a greater degree of home bias and a comparatively low share of external debt.[3]

In the remainder of this paper, we summarize the main insights gained from our historical Greek expedition. Section I presents a brief conceptual discussion on the pitfalls of external financial dependence. In section II, we document that Greece's reliance on foreign savings has been both significant and persistent over the past 200 years; this is evident in the structure of its borrowing, in the country's external position (current account), and in its history of being a large net recipient of foreign grants. In section III, we summarize some dire consequences of Greece's external dependence; we focus on the four episodes of external default (and sudden stops), the protracted crisis resolution in three of these cases, and the heavy political interference from the creditor countries and externally imposed adjustment programs in every case. Section IV addresses the issue of external validity and briefly discusses the relevance of our findings for other countries. In the concluding section, we focus on the current situation and suggest that a significant haircut on the debt stock is needed (that is, on the *external* debt, as sovereign debt is almost entirely in the hands of foreign official creditors).

2. In low-income sub-Saharan Africa, dependence on foreign official financing and aid remains an important challenge.

3. It is an overstatement to conclude that external dependence was not an issue during this period. As we document here, Greece was a major recipient of external aid and grants starting at the end of World War II.

I. External Dependence: Benefits, Costs, and Measurement

Access to external capital markets can deliver many benefits for capital-scarce countries, in particular the possibility to smooth consumption and to use foreign funding for productive investments at home. External debt often carries low interest rates and is readily available, especially in times of high global liquidity. It can therefore be an important complement for more expensive sources of domestic finance.

But these potential advantages of external borrowing may come at a high cost, given the fickle nature of foreign saving.[4] The following risks usually become most apparent during economic crises:

EXTERNAL DEFAULT An obvious first-order risk associated with external debt is that of *external default,* a payment suspension or the restructuring of old debt at terms less favorable to the creditors. Moreover defaults often go hand in hand with (or are the consequence of) a sudden stop in capital flows.

CURRENCY MISMATCH A second peril of external borrowing is rooted in the *currency mismatch* between tax revenues, which are typically in domestic currency, and debt servicing in foreign currency. Since debt crises are intimately connected with currency crises, self-reinforcing vicious spirals are commonplace. This balance-sheet effect can take extreme forms, simultaneously setting the stage for deepening sovereign solvency crises and banking crises.

INABILITY TO TAX FOREIGN CURRENCY DEBT A third pitfall is the *inability to "tax" foreign currency debt* and private foreign investors, be it by spurring inflation or via legislation that reduces the de facto debt servicing costs. Typically, the only mechanism to impose burden sharing and extract relief from external creditors is an outright default and subsequent "haircuts" via negotiated restructurings. It is well known that external creditors (including official ones) typically resist this outcome for as long as possible. In contrast, the government has more options to extract relief when the debt stock is domestic (Reinhart and Sbrancia 2015).

ASYMMETRIC CRISIS SHOCKS A fourth, more subtle, risk of external dependence is the fact that the *"crisis shocks" are asymmetric* if debtors and creditors are not from the same country. Domestic creditors have a strong interest in quick crisis resolution, since they also bear the consequences of a protracted economic downturn that may erode both their income and their wealth. This is not the case for foreign creditors who have less "skin

4. The term "foreign saving" is used interchangeably with "capital inflows."

in the game" when a country enters a severe crisis. Governments also have a much harder time applying regulatory pressure and moral suasion on foreign creditors. This may be one reason why external defaults have tended to last longer than domestic ones, as documented by Reinhart and Rogoff (2009).

EXTERNAL POLITICAL INTERFERENCE Fifth, external borrowing during a boom often ends with heavy *external political interference* during and after the debt crisis. The most drastic examples are military interventions by creditor governments, as in the case of Venezuela in 1902. Less martial but nonetheless powerful forms of foreign interference include the conditionality attached to the granting of rescue loans, as well as conditional aid flows. Political demands in exchange for debt relief have been another vehicle. History is filled with countless examples of creditor governments taking advantage of foreign debt overhang situations as a vehicle to pursue their strategic and economic interests abroad. Arguably, the recent developments in the eurozone crisis are a modern manifestation of foreign interventionism.

To provide a broad picture on external dependence we study indicators of external financial liabilities, sources of government revenue, and proxies for macroeconomic imbalances. In particular, we focus on the level and composition of debt (internal and external, public and private), the current account, transfers and grants from abroad, the "inflation tax," and the scope of domestic savings. We also look at external political pressures and zoom in on changes in external dependence before and after crisis episodes.

On measurement, it is important to note that the lines between what is considered domestic and external debt have become more fluid in the recent wave of financial globalization, largely post-1980s or 1990s. Historically, external debt was issued under foreign law, denominated in a foreign currency (usually the creditor's), and held by nonresidents. Conversely, domestic debt was an entirely domestic affair. In the modern context, as we shall see in the case of Greece, what is domestic in terms of currency or governing law need not be domestic if we look at who actually holds the debt.

II. Greece's Dependence on External Savings: A 200-Year Overview

In this section, we examine Greece's past and present experience with economic crisis, debt, and default, in light of the previous discussion on external dependence.

II.A. Data Preliminaries

Expanding on earlier work by Reinhart and Rogoff (2009) and Reinhart (2010), we begin by constructing a long time series of Greek government debt, breaking it down into its domestic and external components, and dating all external credit events (defaults and restructurings) since Greece's independence. Second, we collected bond-by-bond issuance data using historical investment reports such as Moody's yearbooks, Fenn's compendiums, Kimber's records, the World Bank (Huang and deBeaufort 1954), and the reports by creditor organizations of the time, in particular, the London-based Corporation of Foreign Bondholders (CFB) and the U.S.-based Foreign Bondholder Protective Council (FBPC). The data coverage on Greek bonds is both extensive and well documented, with considerable overlap across sources, resulting in a fairly complete picture of gross borrowing from the rest of the world for the period 1824–1940. In a third step, we gathered data on Greece's current account (from the 1920s onward), private external debt, domestic saving, and post–World War II foreign aid flows, as well as the details of Greece's recent sovereign borrowing. Data on the sources and composition of government revenues and expenditures, inflation, exchange rates, output, and the monetary aggregates (to estimate revenues from the inflation tax) span the 1830s to the present.

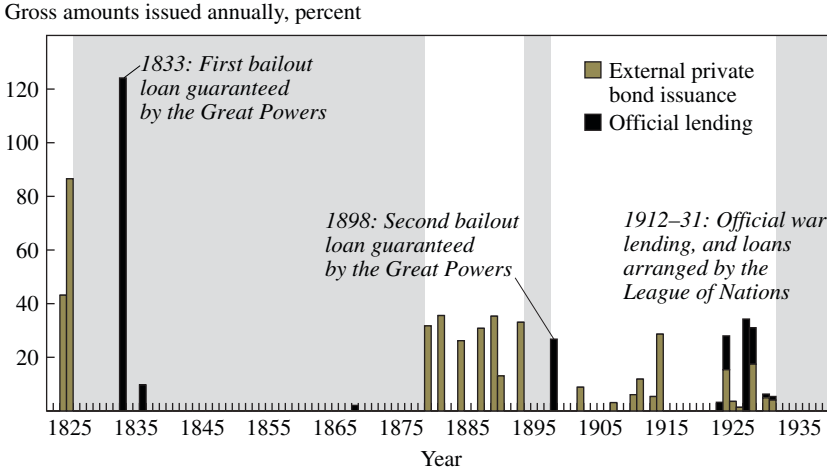
II.B. External Debt

The main insight emerging from this archival work is that Greece has always relied heavily on external borrowing. This can be seen in figure 1, which shows gross external borrowing amounts as a percent of GDP for each year between Greece's War of Independence of the 1820s until World War II. The shaded areas indicate years in default. Lending was mostly from private foreign investors in London and New York (indicated by the bars with light shading). However, during crisis times, the government also became a large-scale borrower from official lenders, in particular, from foreign governments (indicated by the bars with dark shading).

Two main borrowing booms stand out. The first are the very large loans of 1824 and 1825, which were raised in London to finance the independence war against the Ottoman Empire. They imply that Greece started off with an indebtedness above 100 percent of GDP even before gaining independence.[5] The second lending boom occurred after the crisis exit in 1878

5. If the London loans of the 1820s (which were already in default) were combined with the loan of 1833, total indebtedness would have exceeded 200 percent of GDP in that year.

Figure 1. Greek External Borrowing as a Percent of GDP, 1822–1941[a]



Sources: Meyer, Trebesch, and Reinhart (2015) and sources cited therein, in particular, annual reports of the Corporation of Foreign Bondholders, the 1933 Economic Yearbook of Greece, Moody's Manual of Investments, Huang and deBeaufort (1954), and Kimber's records.

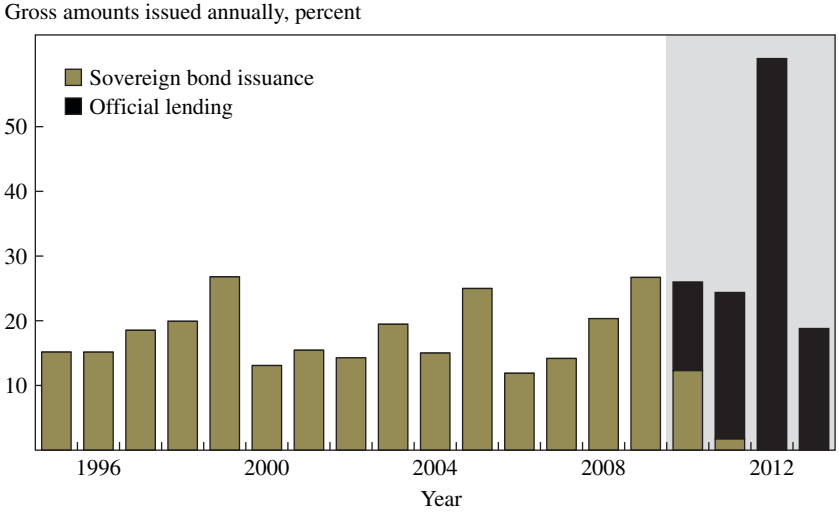
a. Shaded areas represent periods of external default.

and continued until the renewed default in 1893. Within a decade, Greece borrowed more than 100 percent of its GDP from abroad. Once private markets closed, the country continued to borrow from official sources, thus replacing debt on private balance sheets with government-to-government loans.

In recent decades, the borrowing patterns look strikingly similar to the historical picture. Figure 2 summarizes gross sovereign borrowing between 1995 and 2013, using data from Dealogic, Bloomberg, and the European Commission. Sovereign bond issuance in private markets often exceeded 20 percent of the debt-to-GDP ratio annually between 1995 and 2007, and the debt-to-GDP ratio remained at 100 percent, despite high rates of economic growth. After 2010, Greece lost access to private bond markets and again turned to the official sector, with eurozone rescue loans almost entirely substituting for the bonds held by private creditors.

A difference between figures 1 and 2 is that much of the sovereign borrowing in recent decades has been issued under domestic (Greek) law and in domestic currency. Indeed, data by Reinhart and Rogoff (2009) show that the share of domestic borrowing in total Greek sovereign borrowing sees a strong increase after World War II. The picture changes for the 1990s and

Figure 2. Greek Sovereign Borrowing as a Percent of GDP, 1995–2013[a]



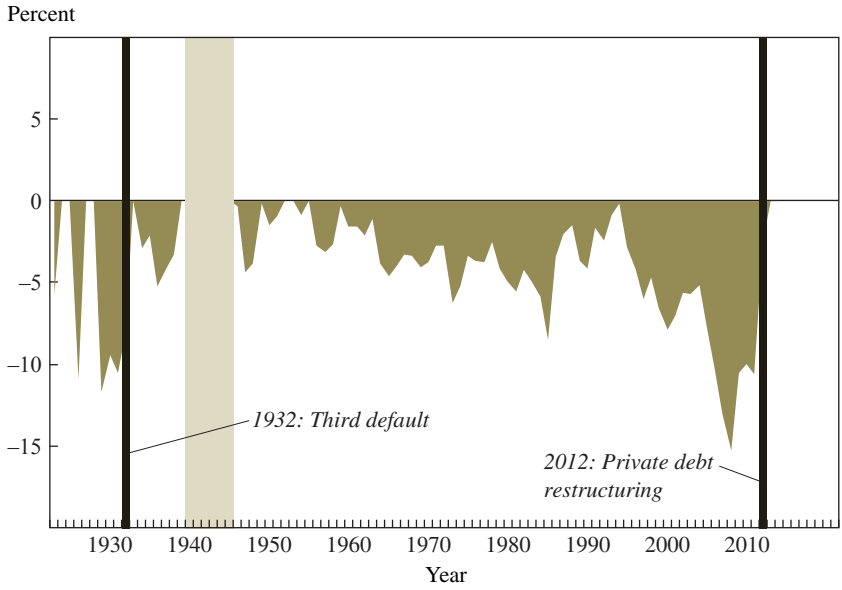
Sources: Bloomberg, Dealogic, and table 9 of European Commission (2014).
a. The shaded area represents the period of debt crisis.

2000s once we measure domestic debt based on who holds the debt—that is, when looking at the creditor base. A significant share of what appears to be domestic debt is actually external by this measure. This can be seen in appendix figure A1,[6] which shows the share of sovereign bonds held by nonresidents. It is remarkable that the share of Greek bonds in the hands of domestic holders has declined from above 70 percent in the late 1990s to about 30 percent prior to the crisis of 2009. Part of this drop can be attributed to a general trend, but Greece shows a much more pronounced decline than other advanced economies and other eurozone periphery countries such as Italy, Portugal, or Spain. This result is corroborated in figure A2 in the online appendix, which shows that since Greece joined the eurozone, the country more than doubled its level of external indebtedness from about 75 percent of GDP in 2001 to 180 percent in 2010.

In sum, the recent boom in borrowing has many features in common with previous Greek surges in borrowing, both in the high levels of debt-to-GDP

6. Online appendixes for papers in this volume may be found at the *Brookings Papers* web page, www.brookings.edu/bpea, under "Past Editions."

Figure 3. Greek Current Account Deficits as a Percent of GDP, 1923–2014[a]



Sources: International Historical Statistics (book series by B. R. Mitchell), Makrydakis (1999), International Monetary Fund (International Financial Statistics and World Economic Outlook), and Reinhart, Reinhart, and Trebesch (2016).

a. No data exist for the World War II period, 1940–45.

that came with it, but also because much of the debt was owed to external creditors. It is striking that each of these *external* debt booms in Greece ended in a painful bust and default, be it in the 1820s, 1880s, 1920s, or 2000s (see appendix figure A3).

II.C. Current Account, Savings, and Grants

Figure 3 takes a different perspective to show that Greece has been, and continues to be, heavily dependent on external savings and highly vulnerable to sudden stops. The country has run current account deficits for almost every year since the 1920s. More precisely, between 1946 and 2014, Greece was in deficit 93 percent of the time, compared to roughly 56 percent of the time for the 19 other advanced economies for which we have data from Reinhart, Vincent Reinhart, and Christoph Trebesch (2016) and sources listed therein. This difference is striking and highly statistically significant.

Moreover, the country has had comparatively low and declining domestic savings, despite the Greek "growth miracle" of the 1960s and 1970s

(appendix figure A4). The savings rate has seen a further drastic collapse since 2008. It is well understood in policy circles, but difficult to quantify, that part of the weakness in domestic saving and the reliance on external saving stems from the fact that much of Greek wealth is held abroad. It is a more or less chronic form of capital flight that intensifies in bad times but is usually present.

Another reason for the enduring current account deficits in figure 3 are grants (as opposed to loans). The country was a net recipient of large-scale aid transfers over much of the post–World War II period, first from the United States, which provided Marshall Plan aid in excess of 5 percent of yearly GDP in the 1950s, and later from the European Economic Community, which transferred yearly grants of 5 percent of GDP after Greece's entry in 1981 (see figure 4).

III. Four Costly Defaults

In the preceding section we have documented how Greece has relied heavily on external savings throughout its history. In this section, we document that this external dependence had a costly downside, particularly in times of crisis.

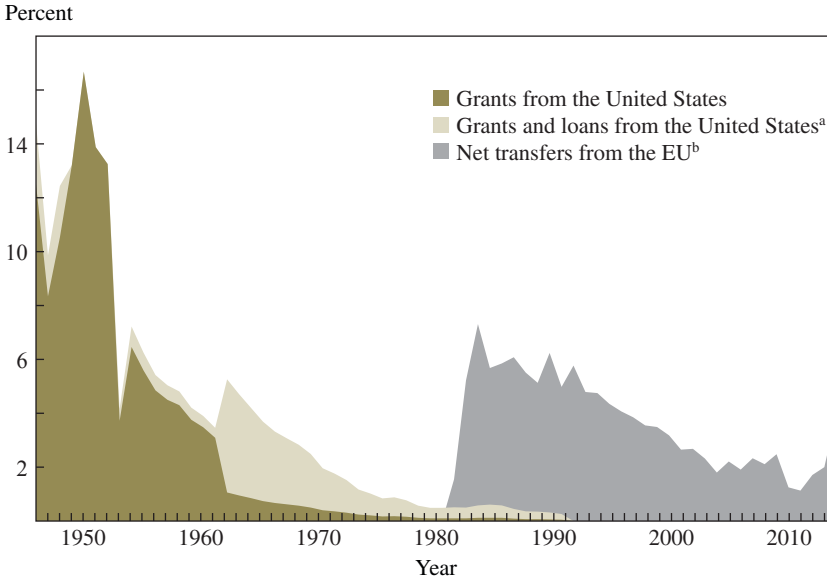
III.A. Repeated Default and Sudden Stops

External debt build-ups in Greece ended in four episodes of external default and sudden stops.[7] In total, the country has been in a state of external default about 50 percent of the years since independence. This can best be seen in table 1, which shows a timeline of main crisis events in the modern history of Greece.

In the run-up to all four debt crisis episodes, Greece lost access to external borrowing and faced increasing interest rates, typical features of a sudden stop. We also find strong balance sheet effects, in particular during the

7. It is five episodes if one treats the default on the guaranteed loan in 1843 as a separate event. It is six episodes if one adds the July 2015 short-lived default on the International Monetary Fund. Other such short-lived defaults are routinely included in the sovereign credit histories documented by the rating agencies (which focus on defaults on private creditors). Moreover, Psalidopoulos and Schönhärl (2012) argue that the first episode of default (from 1826 until 1878) actually includes three separate default events: in 1829, 1836, and 1843, with partial debt servicing occurring between these years. This would increase the default tally to seven. Here, however, we prefer to follow the standard definition of default as periods with missed payments and therefore code the 1826–78 spell as one event. Throughout this period, parts or all of Greece's external debt were not being serviced.

Figure 4. Grants and Loans from the United States and the European Union as a Percent of GDP, 1946–2013



Sources: International Monetary Fund (World Economic Outlook), Maddison Historical GDP Data, Total Economy Database, International Historical Statistics (book series by B. R. Mitchell), USAID (1990), and Bank of Greece.

a. Includes economic and military assistance under various programs, including the postwar relief period (1946–48), the Marshall Plan (1949–51), the Mutual Security Act period (1953–61), and the Foreign Assistance Act period (1962–90).

b. Includes receipts from the European Social Fund (since 1958), the European Agricultural Guidance and Guarantee Fund (since 1962), the European Regional Development Fund (since 1975), and the Cohesion Fund (since 1992).

debt crisis of the early 1930s, in which a drop in the drachma exchange rate and declining central bank reserves resulted in a lack of foreign exchange. The expected exit from the interwar gold standard in 1932 implied that the debt borrowed in dollars and pounds could no longer be serviced out of the state's drachma tax revenues. This contributed to the decision to default in the same months as the gold standard "Grexit" of 1932. Further details on the context of each default episode are provided in appendix C.

III.B. Protracted Crisis Resolution and Limited Debt Relief

Table 1 shows how protracted the resolution of sovereign defaults has been in Greece. The first default of 1826 spanned a remarkable 53 years, while the third default of 1932 was resolved only 30 years later (in 1964).

Table 1. Timeline of Greek Defaults, Bailouts, and External Intervention[a]

Episodes of Default/Debt Crisis		Bailouts and External Interventions
1824/25	Uprising against Ottomans; two loans issued in London to finance war	
1826	Default on the "independence loans" (Debt/GDP > 100%)	
1829	Independence	
1833	King Otto of Bavaria enthroned as King of Greece Guaranteed loan by Great Powers	
1843	Economic crisis and revolt against Otto	Guaranteed loan of 1833 gives **Great Britain, France and Russia** legal control over Greek revenues; high taxes and expense cuts cause public discontent
1862	King Otto overthrown	
1866	Beginning of debt renegotiations	
1878	Debt restructuring and crisis exit	
1879	Market re-access and start of lending boom	
1893	Second default	**1898–1942 International Finance Commission** manages Greek budget and assures debt servicing; financial control imposed by creditor countries as a condition for 1898 guaranteed loan and as part of peace treaty with Turkey
1897	Debt restructuring and peace treaty with Turkey	
1898	Second guaranteed loan by Great Powers	
1912	War lending starts (wars against Turkey and Bulgaria)	
1923	Refugee crisis, loans arranged by League of Nations	
1928	Additional "League Loans"	**1923–32 League of Nations** demands adjustment programs as condition for loans
1932	Third default and exit from gold standard	
1936	Metaxas dictatorship (until 1941)	
1941	Occupation by Nazi Germany and Fascist Italy	
1946	Civil war (until 1949)	
1947	Start of Marshall Plan grants and lending by United States	
1954	Beginning of debt renegotiations	
1964	Debt restructuring and market re-access	
1967	Coup d'état; military junta takes power (until 1974)	
1981	Membership in European Economic Community	**2010–today Troika** of the International Monetary Fund, European Commission, and European Central Bank demands primary surpluses and reforms as condition for bailout loans and eurozone membership
2001	Introduction of euro	
2010	Eurozone bailout; loss of market access	
2012	Private debt restructuring	
2015	Default on the International Monetary Fund (temporary) and third bailout	

a. See appendix C for details.

Moreover, the current debt crisis which started in 2010 is still very far from being resolved.

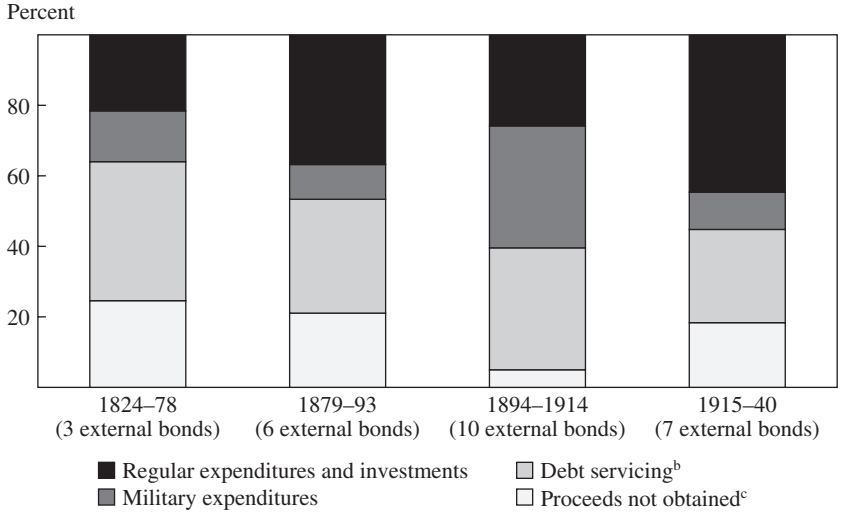
What explains these long delays in crisis resolution in Greece? The reasons are of course manifold, including protracted recessions and the political environments.[8] But part of the delays can be clearly attributed to the creditor side. This is most evident in the largely "excusable" default of 1826 and how it was resolved. We know that the terms of the independence loans of 1824 and 1825 (contracted even before Greece became a sovereign country) were very unfavorable. Of the total nominal value of 2.8 million British pounds borrowed de jure, less than 1.3 million flowed to Greece de facto. The rest were very high commissions and retained amounts due to the issuance price of less than 60 percent of par (see appendix C). In 1829, the government of the newly founded Hellenic Republic approached creditors, offering them to settle the debt so that the repayments would correspond more closely to the actual amounts lent. However, creditors refused to agree to any face-value haircut, and demanded the full repayment of the contractually agreed-upon sums, plus interest payments. With debt above 100 percent of GDP, these demands were difficult to meet in a war-torn and newly founded state.

The refusal to grant debt relief continued after Otto was dethroned and negotiations picked up again. Finally, in 1878, the creditors (or their heirs) agreed to settle the debt at 1.2 million pounds (close to the 1.3 million actually lent) and to forgive the more than 10 million pounds of accrued interest rates and arrears that had accumulated since the 1820s. Ultimately, this restructured debt was then fully repaid upon the pressure of the Great Powers, which exerted a strong influence on Greece in the late 19th century. In other words, the creditor ultimately got back almost the entire nominal amount lent, albeit with a very long delay. The downsides for Greece were 50 years of debt overhang, external interference, and continued exclusion from international markets.

Had the creditors been domestic, the crisis resolution process would most likely have been less protracted, with debt relief granted earlier. The

8. For example, in the reign of King Otto, between 1833 and 1862, Greece refused to even negotiate with creditors, arguing that the war loans of 1824–25 were raised before Greece's independence and therefore not legitimate debt of the Greek state. Moreover, in the early 1950s, Greece underwent a period of heavy political turmoil, with government changing more than once per year. This high turnover rate made it very difficult to engage in long-term negotiations with creditors.

Figure 5. External Bond Proceeds, 1824–1940[a]



Sources: Meyer, Reinhart, and Trebesch (2015) and sources cited therein, in particular the annual reports by the Corporation of Foreign Bondholders and the Foreign Bondholders Protective Council, Levandis (1944), Huang and deBeaufort (1954), and Bikélas (1868).

a. Historically, each sovereign bond prospectus contained a detailed description on how the borrowed amounts would be used. It is therefore possible to categorize the use of proceeds by type.

b. Includes interest and principal on old debt.

c. Includes nominal amounts borrowed that never flowed to the respective country, mostly because the issuance price was often considerably below the par value of the bond.

government would have had more opportunities to pressure domestic holders into an agreement, and domestic creditors might have had more incentives to restructure the debt of their newly independent country. Instead, Greece faced foreign creditors that went out of their way to use their financial and political influence to pressure Greece for repayment, and ultimately largely succeeded in doing so.

Figure 5 illustrates the long-lived consequences of the first external loans of the 1820s and 1830s. The figure breaks down the use of proceeds of each bond borrowed in the first 150 years of Greece's modern history. We separate the share of proceeds that actually benefited Greece's citizens and those that never arrived in the country, either because the new borrowing was used to service old debt or because the issuance price was much below par. The scale of these "non-flows" is striking. Up to the early 20th century, more than 50 percent of the nominal amounts borrowed never arrived in Greece. Moreover, the remaining chunk was then often largely used for military purposes.

Table A1 in the online appendix shows that the use of proceeds did not look more favorably with regard to the bailout loans. It is striking that less than 30 percent of the 1833 guaranteed loan was transferred to the Greek public treasury; the rest was eaten up by fees, retained interest, a large transfer to the Ottoman Empire, and, most importantly, large expenses to install and protect Otto's regency, including the recruitment of a corps of 3,500 Bavarian soldiers. All of this leads L. S. Stavrianos (1952, p. 26) to conclude that "not until 1924 were foreign loans used for productive purposes."

In sum, it took Greece more than 100 years to recover from the legacy of its first external loans. This can also be seen in appendix figure B1, which shows that Greece was running primary surpluses for much of its first 100 years. At the same time, the country was running budget deficits, since the primary surpluses were largely used to service the external debt.

In line with these aggregate numbers, we and Josefin Meyer (2015) calculate that external creditors fared rather benignly in Greece, despite the many years of default. The real ex-post returns on the defaulted bonds were in the range of 1 to 5 percent, despite the losses due to haircuts and arrears. These returns are partly the result of the high yields that these bonds paid between issuance and default, but also because partial debt service continued even in severe crisis years.[9]

Regarding domestic creditors, appendix figure A5 shows that they were heavily taxed, in particular during the interwar years, which saw double-digit inflation. Such "taxation" was never possible with regard to Greece's external creditors. The situation does not look much different today, as Greece enters the fifth year with debt overhang and ongoing discussions on the need for debt relief.

III.C. Foreign Influence, Bailouts, and Recurring Loss of Sovereignty

This last subsection documents how heavy borrowing abroad often resulted in external political dependence. Indeed, we discovered a recurring pattern of bailout lending and related political interference. In each of the four default episodes in Greece, foreign governments stepped in with "rescue" loans, typically in the form of tranches that were conditional on achieving certain fiscal or reform targets. Foreign governments also

9. For example, in the midst of the Great Depression, in 1930 Greece continued to channel more than a third of its budget revenues to service its debt, corresponding to a transfer of 9.25 percent of its GDP, compared to just 2.98 percent in Bulgaria and 2.32 percent in Romania (Stavrianos 1952, p. 26).

succeeded with their demand to impose fiscal and economic policies that assured primary surpluses and a steady flow of debt servicing to private and official creditors abroad.

Table 1 summarizes the episodes of foreign financial control in Greece, while appendix C provides more detailed background information related to each of the four defaults. The first episode resembling a sovereign bailout goes back to 1833, when Great Britain, France, and Russia offered to guarantee a loan raised on private external markets to the ruling King Otto. As collateral for this guaranteed loan, the creditor countries made Greece sign a contract that subordinated all of Greece's revenues, thus giving creditors de jure veto power over Greece's fiscal policies (Kofas 1981; Waibel 2014). This power was exerted most visibly when Greece faced the first major principal payment on these loans but was suffering from an economic downturn. Against the opposition of King Otto and despite the widespread dissatisfaction and protests among the population, the creditor governments demanded full servicing of the 1833 loan, insisting on further budget cuts.

The influence of creditor governments increased further after the renewed default of 1893 and a near defeat in the war against Turkey in 1897. As a condition for arranging a peace treaty with Turkey and in exchange of a new guaranteed loan (that was to be used to pay the war indemnity), the Great Powers, in particular Germany, insisted on establishing an "International Financial Commission," which de facto governed the revenues and expenditures of Greece. The Greek government protested against this loss of sovereignty, but had no choice if it wanted to avoid military defeat. This Commission governed the fiscal policy of Greece for many decades after, until the occupation of Nazi Germany and Fascist Italy ended its rule (Levandis 1944; Waibel 2014).

The scope of external political influence took another turn in the 1920s, when Greece approached the League of Nations to ask for help to tackle the economic downturn and the increasing burden of the refugee crisis from Asia Minor. The League helped to arrange several loans, acting as a trustee. In return, the League negotiated a series of "adjustment programs" with Greece, which were at least partly implemented, in close coordination with the Bank of England, the British Treasury, and the still-powerful Finance Commission (Minoglou 1993).

Against this backdrop, the most recent round of Greek sovereign bailouts and the associated conditionality by the "troika" of the International Monetary Fund, European Central Bank, and European Commission look familiar in regard to the timing, process, and associated political disputes.

As debt migrated from private sector balance sheets to official sector balance sheets, Greece was pushed to give up parts of its sovereignty and to implement adjustment programs to which it did not fully agree.

The success of these interventions was often limited. While the foreign creditors succeeded in enforcing debt repayments over most of the late 19th and early 20th century, the state of Greek finances remained problematic, and the economic conditions unfavorable. In the words of John Levandis (1944, pp. 103–4): "Instead of considering the debt problem in its broad aspects and of adopting measures to eradicate the endemic disease with which Greek finances were perennially afflicted, they introduced half measures of expediency, inadequate to remedy a really difficult and disturbing situation."

Moreover, it is ironic that the crisis resolution with the official sector was no less protracted than that with private creditors. Indeed, as summarized in appendix C, Great Britain, France, and Russia long insisted that the guaranteed loans of 1833 and 1898 were ultimately repaid in full, including any arrears and accrued interest. This resulted in a situation in which Greece was still servicing the bailout debt of 1833 a century later, in the 1930s. As one Greek historian puts it dramatically, "The undeniable fact remains, that the two loans, which were contracted to establish the independence of the Greek state, were the basic factors in its enslavement" (Brewer 2011, p. 296).

Thus, arguably, the most costly legacy of external debt is the loss of political control that comes with it during crisis times.

IV. The Greek Experience in an International Context

Do the pitfalls of external dependence also apply to other countries? Answering this question requires a broad and in-depth analysis, which goes beyond the scope of this paper. But the historical record does indeed suggest that the Greek experience is far from unique.

Maybe the most obvious parallel to Greece is the financial history of Latin America, including countries such as Argentina, Brazil, and Mexico, which have all been chronically dependent on foreign savings and went through repeated boom-bust cycles in international capital flows over much of the past 200 years (Kaminsky and Vega-Garcia 2015). At the same time as being "addicted" to external debt, the region holds the global record in sovereign default years (Reinhart and Rogoff 2009). Moreover, the lost decade of the 1980s debt crisis is also a story of external dependence gone wrong, and shows many resemblances to Greece today, including

large-scale official bailouts, strict adjustment programs, and refusals to grant debt relief by external creditors.

Other examples include Turkey and Egypt, which saw repeated sudden stops and heavy foreign interference in the wake of defaults, as well as some of today's high-income countries such as Portugal, Spain, and China. This latter group featured several lending booms and defaults in the 19th century, but all three countries turned inward in the course of the 20th century, relying more heavily on domestic saving (until recently). Another largely forgotten case is Newfoundland, which lost its sovereignty after defaulting in 1937 (Reinhart and Rogoff 2009).

On the opposite end of the spectrum are countries that have a long history of domestic borrowing, for example Japan, India, and several other Asian countries that have barely witnessed sudden stops and defaults.[10] Moreover, there are countries that successfully "tolerated" large-scale external borrowing from financial centers, in particular Australia, New Zealand, and Canada, which benefited from stable capital inflows even in difficult times (Stone 1999).

V. Conclusion

Sovereign defaults on external creditors can take painfully long to resolve (see table 2). The Greek experience shows that crises can also be very protracted when foreign governments step in and arrange bailout programs, as was the case in the guaranteed Greek loan of 1833. It started out as a loan from private creditors, which Greece could not repay. The 1833 "troika" (the Great Powers of France, Great Britain, and Russia) repaid the private creditors, and Greece's debts shifted to official hands. After decades in default and financial autarky, Greece still faced repayment of that loan more than 100 years later. Such a crisis resolution approach, which results in decades of debt overhang, perpetuates external dependence and impedes a "fresh start" for the over-indebted country.

We have documented elsewhere that protracted debt crises are typically resolved only after creditors agree to face-value haircuts (Reinhart and Trebesch 2016). Decisive debt relief is associated with higher subsequent growth that softer forms of debt relief, such as maturity extensions, do not usually deliver. Against this backdrop, a key ingredient in the resolution to the ongoing Greek crisis is a deep nominal haircut on the stock of official

10. The only Japanese default was in the wake of World War II, and that was on small amounts only; see Meyer, Reinhart, and Trebesch (ongoing work).

Table 2. Elements of Greek Debt Resolution, 1826–2015

Period	*Delay*	*Bonds restructured*	*Interest arrears*	*Haircut*
Default of 1826–78	53 years	2.8 million pounds	10 million pounds	Between 40 percent (face-value reduction) and 91 percent (including cancelation of interest arrears)
Default of 1893–98	5 years	22.3 million pounds	3 million pounds	Between 37 percent and 53 percent, depending on assumptions; no face-value reduction
Default of 1932–64	32 years	54.7 million pounds	64.5 million pounds	Between 64 percent (excluding interest arrears) and 86 percent; no face-value reduction
Debt restructuring of 2011–12	Less than 1 year	199.2 billion euros	None (preemptive)	Between 59 percent and 65 percent, depending on discount rate and assumptions

Sources: Meyer, Reinhart, and Trebesch (2015), Zettelmeyer, Trebesch, and Gulati (2013), and sources cited therein.

(and possibly private) external debt. Further maturity extensions would be an unfortunate repetition of the Greek history documented in this paper (see table 2) and would only delay the day of reckoning—to the disadvantage of both Greek and eurozone taxpayers. Extending the debt until 2070 (as discussed by the International Monetary Fund [2015]) is likely to add fuel to a never-ending debate on what to do with Greek debt. It is difficult to see how this could foster a renewal of confidence and sustained growth and investment.

Beyond the immediate and towering challenge of coping with the current debt crisis, we believe that Greece (and periphery Europe) can learn from some of the measures taken in many emerging markets in the 1990s after their own financial crises. We are well aware that Asia in the 1990s, in particular, started from a much more favorable position than Greece today. Nonetheless, a long-run policy priority for Greece should be to shift the balance to domestic sources of funding. Since the late 1990s, numerous emerging market governments have, in varying degrees, reduced their reliance on external financing by tilting new debt issuance to the domestic

market. Prudential public debt management, however, does not directly address the vulnerabilities posed by surges in private external borrowing. To deal with the macroeconomic risks often connected to the latter, many countries have adopted policies that tax or limit some or all forms of external borrowing or foreign exchange exposure. Whether such policies fall under the broad headings of capital controls or macroprudential regulation has depended on the particulars of each case.

Overall, we have no basis to conclude that greater reliance on domestic savings will be a panacea of economic stability, but we do have 200 years of evidence to support the view that chronic reliance on external capital has repeatedly led to ruin.

ACKNOWLEDGMENTS We wish to thank the Editors, Vassilis Droucopoulos, George Georganas, Josefin Meyer, Michael Papaioannou, Michalis Psalidopoulos, Vincent Reinhart, David Romer, Julian Schumacher, Constantine Yannelis, and Robert Zoellick for helpful comments, and Jochen Andritzky and Stelios Makrydakis for sharing the data from their studies, which we cite here. Sebastian Horn, Maximilian Mandl, and Maximilian Rupps provided invaluable research assistance.

References

Bikélas, Demétrius. 1868. "Statistics of the Kingdom of Greece." *Journal of the Statistical Society of London* 31, no. 3: 265–98.

Brewer, David. 2011. *The Greek War of Independence: The Struggle for Freedom from Ottoman Oppression.* New York: Overlook Press.

Corsetti, Giancarlo, Lars P. Feld, Philip R. Lane, Lucrezia Reichlin, and others. 2015. "A New Start for the Eurozone: Dealing with Debt." Online article. London: Centre for Economic Policy Research, VoxEU. http://www.voxeu.org/article/new-cepr-report-new-start-eurozone-dealing-debt

Eichengreen, Barry, Naeun Jung, Stephen Moch, and Ashoka Mody. 2014. "The Eurozone Crisis: Phoenix Miracle or Lost Decade?" *Journal of Macroeconomics* 39, pt. B: 288–308.

European Commission. 2014. "The Second Economic Adjustment Programme for Greece: Fourth Review." Occasional Paper no. 192. Brussels.

Gros, Daniel. 2013. "Foreign Debt versus Domestic Debt in the Euro Area." *Oxford Review of Economic Policy* 29, no. 3: 502–17.

Huang, Andrew C., and Carel deBeaufort. 1954. "External Public Debt of Greece." Sector Report no. EC28. Washington: World Bank.

International Monetary Fund. 2015. "Greece: An Update of IMF Staff's Preliminary Public Debt Sustainability Analysis." Country Report no. 15/186. Washington.

Kaminsky, Graciela L., and Pablo Vega-Garcia. 2015. "Systemic and Idiosyncratic Sovereign Debt Crises." Working Paper no. 20042. Cambridge, Mass.: National Bureau of Economic Research.

Kofas, Jon V. 1981. *Financial Relations of Greece and the Great Powers, 1832–1862.* New York: East European Monographs.

Levandis, John A. 1944. *The Greek Foreign Debt and the Great Powers, 1821–1898.* Columbia University Press.

Makrydakis, Stelios. 1999. "Consumption-Smoothing and the Excessiveness of Greece's Current Account Deficits." *Empirical Economics* 24, no. 2: 183–209.

Meyer, Josefin, Carmen M. Reinhart, and Christoph Trebesch. Ongoing work. "200 Years of Sovereign Haircuts."

Minoglou, Ioanna Pepelasis. 1993. "The Greek State and the International Financial Community, 1922–1932: Demystifying the Foreign Factor." Ph.D. Thesis, London School of Economics and Political Science. http://etheses.lse.ac.uk/2590/

Psalidopoulos, Michalis, and Korinna Schönhärl. 2012. "Die griechische Staatsverschuldung in der zweiten Hälfte des 19. Jahrhunderts" [The Greek Government Debt in the Second Half of the 19th Century]. In *Vom Wohl und Wehe der Staatsverschuldung: Erscheinungsformen und Sichtweisen von der Antike bis zur Gegenwart* [On the Boon and Burden of the National Debt: Manifestations and Perspectives from Antiquity to the Present], edited by Thorsten Beigel and Georg Eckert. Münster: Aschendorff.

Reinhart, Carmen M. 2010. "This Time Is Different Chartbook: Country Histories on Debt, Default, and Financial Crises." Working Paper no. 15815. Cambridge, Mass.: National Bureau of Economic Research.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. *This Time Is Different: Eight Centuries of Financial Folly.* Princeton University Press.

Reinhart, Carmen M., Vincent R. Reinhart, and Christoph Trebesch. 2016. "Global Capital Flow and Commodity Cycles: A Very Long View." Working Paper.

Reinhart, Carmen M., and M. Belen Sbrancia. 2015. "The Liquidation of Government Debt." *Economic Policy* 30, no. 82: 291–333.

Reinhart, Carmen M., and Christoph Trebesch. 2016. "Sovereign Debt Relief and Its Aftermath." *Journal of the European Economic Association* 14, no. 1: 215–51.

Sinn, Hans-Werner. 2014. *The Euro Trap: On Bursting Bubbles, Budgets, and Beliefs.* Oxford University Press.

Stavrianos, L. S. 1952. *Greece: American Dilemma and Opportunity.* Chicago: Henry Regnery Company.

Stone, Irving. 1999. *The Global Export of Capital from Great Britain, 1865–1914: A Statistical Survey.* New York: St. Martin's Press.

USAID (U.S. Agency for International Development). 1990. "U.S. Overseas Loans and Grants and Assistance from International Organizations." Washington.

Waibel, Michael. 2014. "Echoes of History: The International Financial Commission in Greece." In *A Debt Restructuring Mechanism for Sovereigns: Do We Need a Legal Procedure?,* edited by Christoph G. Paulus. Munich, Oxford, and Baden-Baden: C. H. Beck, Hart, and Nomos.

Zettelmeyer, Jeromin, Christoph Trebesch, and Mitu Gulati. 2013. "The Greek Debt Restructuring: An Autopsy." *Economic Policy* 28, no. 75: 513–69.

CHRISTOPHER L. HOUSE
University of Michigan

LINDA L. TESAR
University of Michigan

Greek Budget Realities: No Easy Option

ABSTRACT This paper uses a quantitative dynamic open economy macroeconomic model to examine alternative strategies that the Greek government could implement to increase its primary balance on a flow basis by 1 percent of GDP, representing roughly one quarter of Greece's total annual liability. We examine the impact of increases in distortionary taxes and reductions in government expenditures on the macroeconomy in both the short and long run. The necessary fiscal adjustments are large and entail substantial macroeconomic costs. These costs are even greater when one takes into account realistic elasticities of the tax base and the fact that Greece is a small open economy. Delaying fiscal adjustment could yield short-term benefits, but ultimately such delays come at a high price unless Greece's creditors are willing to provide additional finance at below-market rates. The basic framework holds the growth rate of the Greek economy fixed. Naturally, fiscal adjustments become less painful under a scenario in which the Greek economy returns to a positive growth path. Whether structural reforms or other policies can generate such growth remains an open question.

As of August 2015, Greece's loan repayments due to external creditors through 2057 summed to €319.5 billion, requiring an average debt payment on a flow basis of 4.1 percent of 2014 Greek GDP.[1] This paper examines the economic impact of alternative strategies that Greece could implement to generate a sizable primary surplus. To be specific, we consider increases in distortionary taxes on consumption, capital, and labor income, as well as reductions in government expenditures sufficient to increase Greece's primary balance by 1 percent of 2014 GDP—roughly

1. This sum reflects a variety of interest rates, revolving credit agreements, and bond maturities. The International Monetary Fund's Dissemination Standards Bulletin Board reports budgetary central government debt of €312.8 billion for the second quarter of 2015.

a quarter of Greece's total debt obligations. By studying each policy in isolation, we illustrate the costs and benefits associated with each. In the discussion in section IV, we also consider combinations of policies that could yield the necessary 4.1 percent.[2]

Our analysis yields the following conclusions:

—In the baseline case calibrated to the Greek economy, all of the tax and expenditure policies that we consider produce declines in output in both the short and the long run. The model projection for the near term involves output declines on the order of 1 to 2 percent of 2014 GDP.

—Projections of the primary surplus based on static revenue scoring grossly overestimate the actual amount of revenue that Greece would raise from tax increases. The overestimate is because the static projections ignore endogenous adjustments of capital and labor.

—Meeting the debt repayment schedule is substantially more costly because Greece is a small economy that is integrated with the larger European economy. Failure to incorporate the impact of capital and labor mobility results in a significant overestimate of future revenue.

—Delaying the implementation of tax increases or government expenditure cuts can help mitigate the short-run fall in output, but such delays require greater economic hardship in the long run.

I. Framework for Analysis

The basic model for our analysis includes features of dynamic stochastic general equilibrium models commonly used by central banks and macroeconomists to study both short-run business cycle dynamics and long-run adjustment to permanent changes in fiscal policy. At its core, the model closely resembles the framework used by Jonathan Heathcote and Fabrizio Perri (2002), in which foreign and domestic intermediate goods are combined to create a "final good" that is ultimately used for domestic consumption, investment, and government purchases.[3] Greece, the home country, is treated as a small part of a larger European aggregate. Greece faces a downward sloping demand curve for its country-specific export good, but the share of Greece in the European economy is so small that events in Greece have no meaningful effect on economic activity in the European aggregate.

2. The Memorandum of Understanding between the European Commission, the Greek government, and the National Bank of Greece, agreed to in August 2015, included a commitment on the part of the Greek government to run primary surpluses from 2018 forward of 3.5 percent of (current) GDP (European Commission 2015).

3. Given space constraints, we do not provide a detailed discussion of the model in this paper. We refer the interested reader to House, Proebsting, and Tesar (2015).

Tradable intermediate goods are produced with capital and labor inputs. Bond markets, asset markets, and currency markets in the euro area are fully integrated, and as a member state, Greece takes the nominal interest rate set by the European Central Bank (ECB) as given. In the baseline model, labor cannot move across international borders, but that assumption of labor immobility is belied by the outflow of labor, particularly skilled workers, from Greece since the onset of the crisis. To capture the effect of labor mobility, we consider high labor supply elasticity as a special case. Not surprisingly, when labor responds more elastically to increases in tax distortions it is much harder to raise revenue, and the excess burden of the tax on the labor that remains in Greece is larger.

The Greek government is assumed to raise revenue through distortionary taxes on capital and labor income and on consumption spending. Part of the revenue is transferred back to households, part is used to finance government spending on final goods, and some is transferred to foreign creditors as debt repayment. We assume that government purchases do not affect the marginal utility of consumption, the marginal disutility of labor, or production. The model includes sluggish adjustment in prices and wages and allows for adjustment costs in investment.

We examine the impact on macroeconomic aggregates in response to once-and-for-all changes in policy over different time intervals and under different specifications of model parameters. Our analysis can be thought of as an extension of the tax experiments considered by Greg Mankiw and Matthew Weinzierl (2006), where here we include integrated capital markets and a richer set of adjustment mechanisms, and examine changes in macroeconomic variables along the transition to the post-policy steady state.

I.A. Distortionary Labor and Consumption Taxation

Labor and consumption taxation jointly reduce the incentive to work, depress the supply of labor, and reduce consumption. In the absence of wage rigidity, these distortions would enter through the labor supply condition

$$v'(N_t) = \frac{W_t(1-\tau_t^N)}{P_t(1+\tau_t^C)}u'(C_t). \tag{1}$$

N_t denotes total hours of labor supply and $v'(N_t)$ is the marginal disutility of labor; C_t is real consumption and $u'(C_t)$ is the marginal utility of consumption. The term $W_t(1-\tau_t^N)/P_t(1+\tau_t^C)$ is the after-tax real wage. Increases in the tax rate on labor income τ_t^N or the tax rate on consumption spending τ_t^C reduce labor supply by reducing the real after-tax

compensation for work. For the numerical model, we assume the flow utility function takes the form

$$u(C_t) - v(N_t) = C_t^{1-\frac{1}{\sigma}} - \phi N_t^{1+\frac{1}{\eta}}.$$

The parameter η is the Frisch labor supply elasticity; higher values of η imply that workers are more responsive to changes in the after-tax real wage. The parameter σ is the intertemporal elasticity of substitution. While the consumption-labor supply choice is distorted by the wedge that jointly reflects the labor and consumption tax, the taxes are not equivalent in their revenue implications, and therefore we consider each tax separately.

The labor supply condition (equation 1) embodies a second channel through which changes in government finance influence labor supply. A reduction in consumption, for example one triggered by an increase in government spending, results in an increase in labor supply through a wealth effect. Events or policies that reduce wealth and thus reduce consumption are also predicted to ultimately raise employment by giving workers a greater incentive to work. (This is sometimes referred to as an "impoverishment effect" on labor supply.)

The model we analyze includes both wage and price rigidity, and as a consequence the simple labor supply condition described in equation 1 does not hold at every moment, though the basic economic intuition discussed above remains intact. To capture the slow adjustment of nominal prices and wages, the model includes Calvo price- and wage-setting mechanisms, following Lawrence Christiano, Martin Eichenbaum, and Charles Evans (2005). The rates at which prices and wages adjust are governed by two Calvo parameters. The Calvo mechanism allows for fully flexible prices and wages as a special case.

I.B. Distortionary Capital Taxation

Capital taxes reduce the after-tax payoff to capital accumulation and thereby reduce the incentive to expand business activity. This effect typically takes some time to materialize. In the short run, the capital stock is already in place, and thus the supply of capital is relatively inelastic. In the model, this distortion enters through the investment demand schedule, which itself is a combination of the two equations

$$Q_t = \beta E_t \left[\sum_{j=1}^{\infty} \beta^j (1-\delta)^j \frac{u'(C_{t+j})}{u'(C_t)} \frac{R_{t+j}(1-\tau^K_{t+j})}{P_{t+j}} \right] \tag{2}$$

and

$$\Delta \ln I_t = \frac{1}{\kappa} \ln Q_t + \beta E_t [\Delta \ln I_{t+1}]. \tag{3}$$

Equation 2 gives the relative value of capital (Tobin's Q) in terms of the discounted stream of real capital payoffs. In equation 2, $R_t(1 - \tau_t^K)/P_t$ is the real after-tax flow payoff to a unit of capital. Equation 3 provides a connection between the relative value of capital (Q_t) and investment. We adopt the approach in Christiano, Eichenbaum, and Evans (2005), who use "higher order adjustment costs" to generate additional persistence in investment. The degree of adjustment costs is then governed by a single parameter $\kappa \geq 0$. Expected increases in future capital taxes reduce the relative value of capital through equation 2, and that in turn depresses current investment.

In a closed economy, an increase in the capital tax rate would reduce capital investment, but some of the negative impact of the tax increase would be offset by a change in the domestic interest rate. Because Greece is part of a global financial system, Greek savers have the option of substituting away from investing in domestic capital to investing in international bonds. This channel of substitution will increase the long-run elasticity of capital with respect to the tax rate.[4]

I.C. Government Solvency Constraint

In all of the model simulations, we assume that Greece chooses a policy combination that enables it to repay a given amount of money to its foreign creditors. This requires that the policy options satisfy the government's budget constraint

$$\sum_{j=1}^{\infty} \left(\frac{1}{1+i} \right)^j [P_t G_t + T_t + X_t - \tau_t^K R_t K_t - \tau_t^N W_t N_t - \tau_t^C P_t C_t] = 0, \tag{4}$$

where X_t is a nominal repayment commitment, T_t is a nominal lump sum transfer (or tax), and $i_t = i$ is the nominal interest rate.

4. Because the final investment good is country-specific, physical capital is immobile in the sense that a unit of capital cannot be unbolted from the home factory and put into operation abroad. However, the capital stock does gradually adjust to changes in the tax rate through the substitution between investment in new domestic capital and international bonds.

We take 2014 as the point of departure and consider once-and-for-all changes to economic policy starting in 2015. Our baseline analysis assumes that the debt repayment commitment is not anticipated prior to 2014—that is, we assume that the indebtedness comes as a surprise to the Greek population.[5] Later, we relax this assumption and consider the case where the fiscal policy changes occur with a lag so that households anticipate future changes in policy. In all cases, we assume that Greece maintains access to financial markets (that is, it has sufficient credibility with its creditors) to borrow in years when debt payments exceed the fiscal surplus, and saves when the surplus exceeds its scheduled debt payment.[6]

To facilitate comparison across policy experiments, we consider spending reductions or tax increases sufficient to generate an average flow increase in the primary balance of 1 percent of 2014 GDP. For ease of exposition, we will refer to the permanent 1 percent increase in the primary balance as the *target* primary balance. The target we consider is only a quarter of the amount required to fully meet the scheduled stream of debt payments. Nonetheless, the magnitudes of the tax and expenditure changes needed to generate the target primary balance are already quite large. We do not push the model to generate the full 4 percent increase in the primary balance as a share of 2014 GDP for three reasons. First, to solve the model we use standard linearization techniques, which are appropriate for "small" perturbations in the neighborhood of a stable-trend growth path but are increasingly inaccurate for large policy changes that push the economy away from the initial steady state. In principle, a tax policy change that would satisfy (or attempt to satisfy) the full 4.1 percent increase could push capital and labor taxes into the downward-sloping portion of the Laffer curve (Trabandt and Uhlig 2006; Mendoza, Tesar, and Zhang 2014). Second, our baseline revenue forecasts are based on a model with zero economic growth. This calibration is in line with both recent experience in Greece as well as forecasts of long-run growth absent significant structural reform (McQuinn and Whelen 2015). Were the Greek economy to return to a positive rate of growth, the policy changes needed to meet the target would be smaller. We

5. To be precise, we assume that the Greek economy begins in an initial steady state. Then, in 2014, Greek taxpayers simultaneously learn that their debt obligations X_t are larger than previously believed and learn about the new policies aimed at reducing the debt.

6. In reality Greece faces a lumpy stream of payments, with large payments due in some years and smaller payments in others. An interesting question we leave for future study is the case when Greece faces binding credit constraints and must either accumulate balances in advance of large payments or adjust fiscal policy on a year-by-year basis.

return to the issue of economic growth later in the paper. Finally, the actual policy response is likely to be a combination of tax increases and spending cuts, rather than an increase in a single policy instrument.

II. Calibration

Table 1 lists the parameters used to solve the model along with their economic interpretation. The rate of time discount $\beta = 0.99$, and the intertemporal elasticity of substitution $\sigma = 0.50$. In the baseline case, the elasticity of substitution between home and foreign goods $\psi = 1.50$ (constant elasticity of substitution preferences over home and foreign goods). We also consider the case in which home and foreign goods are perfect substitutes (the neoclassical model in table 2). A key parameter in the model is the Frisch labor supply elasticity η, which we set at $\eta = 0.5$ in the baseline model (Chetty and others 2011). To proxy for international labor mobility, we assume a very high Frisch elasticity of $\eta = 10$ in one of the model specifications.

Turning to the production side of the model, we assume that the production function is Cobb-Douglas with a labor share $\alpha = 0.64$. We set adjustment costs on investment at roughly $\kappa = 2.5$, implying that a 1 percent increase in Tobin's Q causes a 0.4 percent increase in investment growth, all else equal. We set the Calvo parameters so that wages adjust on average once every 12 months (Barattieri, Basu, and Gottschalk 2014) and prices adjust on average once every 8 months (Nakamura and Steinsson 2008; Klenow and Kryvtsov 2008). Greece's share of total GDP in the system is 2 percent, consistent with Greece's GDP share of Europe. The share of imports relative to GDP is 0.21, consistent with the average observed import share from 2000 to 2013.

The initial stance of fiscal policy reflects the level of taxes and government expenditures observed in Greece in 2014 (or the latest year for which data are available). Estimates of average effective tax rates τ on labor, capital, and consumption are calculated using the methodology developed by Enrique Mendoza, Assaf Razin, and Linda Tesar (1994); they are 43, 18, and 16 percent, respectively. The bottom section of table 1 shows macro aggregates as shares of 2014 GDP. The consumption and investment shares are determined by the model while the other ratios are preset. We start from a benchmark trade-balance-to-GDP ratio of balanced trade, ignoring Greece's large initial current account deficit. We also set the primary-balance-to-GDP ratio equal to zero, consistent with the data from 2014. Government spending as a share of goods and services is set at 20 percent, slightly higher than the rate observed in 2014 (18 percent). The

Table 1. Parameters and Calibration

Preferences		
Subjective time discount factor, β		0.99
Trade elasticity, ψ		1.50
Intertemporal elasticity of substitution, σ		0.50
Frisch elasticity, η		0.5, 10
Technology		
Labor share, α		0.64
Inverse Q elasticity, κ		2.48
Duration of price rigidity		8 months
Duration of wage rigidity		12 months
Desired price and wage markup		0.10
Greek share of EU GDP		0.02
Greek import share		0.21
Initial effective tax rates	**Data**	**Model**
Labor income tax rate, τ^N	0.43	0.43
Capital income tax rate, τ^K	0.18	0.18
Consumption tax rate, τ^C	0.16	0.16
Ratios to GDP in 2014[a]		
Consumption	0.74	0.61
Investment	0.13	0.19
Government expenditures	0.18	0.20
Trade balance	–0.08	0.00
Primary balance	0.00	0.00

a. Empirical ratios based on 2014 (or most recent year available) data from Eurostat and the Organization for Economic Cooperation and Development.

model generates a consumption share that is lower than that in the data and an investment share that is too high, largely because it ignores the current account imbalance that supports higher Greek consumption.

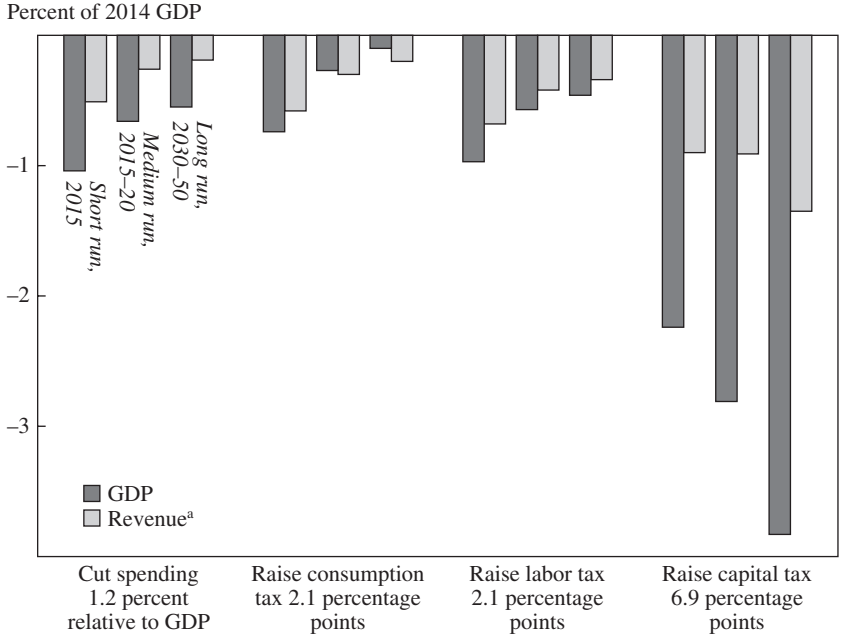
III. Results

Figure 1 illustrates the change in GDP (dark gray) and the difference between the actual and static revenue scores (light gray) associated with the various policy changes needed to reach the target primary balance in the baseline model. The figure shows the effect of the policy in the short run (2015), the medium run (2015–20), and the long run (2030–50). Table 2 reports results for output and employment for each policy change and compares results across different model specifications.

III.A. Decrease in Government Spending

In the baseline model, a decrease in government spending (the first set of bars) results in a roughly 1 percent decline in GDP in the short run,

Figure 1. Effects of Policies on GDP and Revenue Scores



a. Actual revenue score minus static revenue score.

with slightly smaller average declines over the medium and long runs. If the policymaker had relied on a static score to estimate the impact of the policy, he or she would have assumed that there was no effect of the expenditure cut on revenue, and would therefore have underestimated the size of the expenditure cut needed to meet the repayment target. In the dynamic model, however, the decrease in government expenditures reduces aggregate demand for the home good, resulting in declines in labor and capital income as well as consumption and therefore in a decline in tax revenue. The gap between the static score and the dynamic score is roughly 0.5 percent of 2014 GDP.

If the economy is truly neoclassical (see table 2, second panel), with flexible prices and wages and perfect substitutability between the home and the foreign good (essentially a one-good economy), the reduction in government spending has no impact on macroeconomic activity beyond the transfer of domestic resources to foreign creditors. In this case, the government reduces its purchases by exactly enough to repay its foreign creditors. Net exports rise by exactly the amount that government demand falls, leaving aggregate demand (and thus economic activity)

Table 2. Effects of Policies on GDP, Employment, and Revenue Scores

	Government spending cut			*Consumption tax increase*		
	Short run, 2015	*Medium run, 2015–20*	*Long run, 2030–50*	*Short run, 2015*	*Medium run, 2015–20*	*Long run, 2030–50*
Baseline model						
	Cut spending by 1.21 percent of GDP			Tax increase from 16 to 18.2 percent		
GDP	–1.04	–0.66	–0.55	–0.74	–0.27	–0.10
Employment	–1.05	–0.23	0.06	–0.60	0.34	0.62
Static revenue	0.00	0.00	0.00	1.22	1.22	1.22
Actual revenue	–0.51	–0.26	–0.19	0.64	0.92	1.02
Neoclassical model						
	Cut spending by 1 percent of GDP			Tax increase from 16 to 17.7 percent		
GDP	0.00	0.00	0.00	0.27	0.34	0.47
Employment	0.00	0.00	0.00	0.40	0.43	0.47
Static revenue	0.00	0.00	0.00	0.98	0.98	0.98
Actual revenue	0.00	0.00	0.00	0.95	0.97	1.01
Sticky prices and wages						
	Cut spending by 1 percent of GDP			Tax increase from 16 to 17.7 percent		
GDP	–0.01	0.00	0.00	0.21	0.33	0.47
Employment	–0.01	0.00	0.00	0.31	0.41	0.47
Static revenue	0.00	0.00	0.00	0.98	0.98	0.98
Actual revenue	0.00	0.00	0.00	0.93	0.97	1.01
International labor mobility						
	Cut spending by 1.3 percent of GDP			Tax increase from 16 to 18.7 percent		
GDP	–1.50	–1.08	–0.81	–1.45	–1.06	–0.80
Employment	–1.84	–0.84	–0.24	–1.75	–0.81	–0.23
Static revenue	0.00	0.00	0.00	1.51	1.51	1.51
Actual revenue	–0.63	–0.39	–0.28	0.70	0.92	1.02

unchanged. This is not to say that the Greek citizens do not care about this policy change; while GDP and employment are unaffected, welfare decreases to the extent that households value the government services that are no longer being provided.

III.B. Increases in Labor and Consumption Taxes

Consumption and labor tax rates jointly affect the wedge between the marginal rate of substitution and the marginal product of labor, as seen in equation 1. In the baseline model, the increase in the consumption and labor tax rates is about the same size, a 2.1 percentage point increase. The static score overstates the increase in revenue by roughly 0.3 percent of 2014 GDP.

Labor tax increase			*Capital tax increase*		
Short run, 2015	*Medium run, 2015–20*	*Long run, 2030–50*	*Short run, 2015*	*Medium run, 2015–20*	*Long run, 2030–50*
Tax increase from 43 to 45 percent			Tax increase from 18 to 25 percent		
–0.97	–0.57	0.46	–2.24	–2.81	–3.83
–0.96	–0.10	0.18	–1.74	–0.80	0.51
1.36	1.36	1.36	2.24	2.24	2.24
0.68	0.94	1.01	1.34	1.33	0.89
Tax increase from 43 to 44.8 percent			Tax increase from 18 to 26.6 percent		
0.06	0.07	0.10	0.73	–1.70	–6.13
0.09	0.09	0.10	1.98	1.17	–0.24
1.12	1.12	1.12	2.79	2.79	2.79
0.99	0.99	1.00	2.86	2.06	0.77
Tax increase from 43 to 44.8 percent			Tax increase from 18 to 26.6 percent		
0.04	0.07	0.11	0.37	–1.79	–6.14
0.06	0.09	0.11	1.44	1.05	–0.23
1.12	1.12	0.12	2.80	2.80	2.80
0.98	0.99	1.00	2.77	2.04	0.77
Tax increase from 43 to 47.3 percent			Tax increase from 18 to 31.1 percent		
–3.12	–3.77	–4.06	–5.26	–8.45	–9.64
–4.35	–4.79	–4.20	–5.82	–6.92	–2.32
2.78	2.78	2.78	4.24	4.24	4.24
1.28	1.13	0.97	2.64	1.71	0.81

The responses of labor and output to the increase in taxes depend on the relative strength of the income and substitution effects on labor supply and on the elasticity of substitution between domestic and foreign goods. For both the neoclassical model and the sticky price model, the income and substitution effects on labor supply nearly offset one another. The tax increase makes households poorer and increases their incentive to work. In our model, the income elasticity of labor supply is given by $-\eta/\sigma$, while the (Frisch) substitution elasticity is η. Since $\sigma = 0.5$ in our baseline calibration, the income effect dominates the substitution effect, so labor supply and output ultimately increase. The importance and realism of the wealth effect are somewhat controversial. Richard Blundell and Thomas MaCurdy

(1999) report empirical estimates of income elasticities of labor supply between 0 and −1, with typical estimates in the range of −0.1 to −0.2. Our calibration implies an income elasticity of −1, the top of the range reported by Blundell and MaCurdy (1999). Reducing the income elasticity would imply larger short-run output drops than those shown in figure 1.

In addition to the income and substitution effects on labor supply, the equilibrium also depends heavily on the trade elasticity. In both the neoclassical and the sticky price specifications, this elasticity is infinite. As a consequence, Greece can pay off its nominal obligations simply by working more and selling home goods to its foreign creditors at par. In the baseline specification, however, the trade elasticity is only 1.5. This implies that as the supply of Greek goods rises, the price of these goods falls, tempering the payoff to working more. In the baseline model, this trade effect causes employment and output to fall in equilibrium.

III.C. Increase in the Capital Tax

The capital tax is the most inefficient in terms of generating revenue per unit of output loss. It also generates the largest gap between actual revenue and the static projection. An increase in the capital tax leads directly (subject to adjustment costs) to a reduction in investment and an associated decline in output and employment. The erosion of the tax base is amplified by households' ability to substitute away from investment in the home capital good toward investment in the international bond. This erosion of the tax base explains why a much larger tax increase (6.9 percentage points) is needed to meet the revenue target. In the baseline model as well as the neoclassical model, an increase in the capital tax causes capital, labor, and output to fall in the long run.

III.D. Adding Sticky Prices to the Neoclassical Model

The third panel of table 2 reports results for a "sticky price" specification. This specification is identical to the neoclassical specification with the exception that prices and wages adjust slowly to changes in economic activity, as they do in our baseline model. Importantly, the sticky price specification maintains the assumption that there is an infinite elasticity of substitution between home and foreign goods, so this is effectively still a one-good model.

Price and wage rigidity have only modest effects relative to the predictions of the basic neoclassical model. Virtually all of the differences between the two model specifications are concentrated in the short-run responses, and even these differences are relatively small. In the long run, the two

models deliver essentially identical predictions. Prices and wages eventually adjust to the policy change, at which point the model has returned to the same trajectory as the one for an environment with fully flexible prices.

The responses to the government spending cut are identical with and without sticky prices. Since there was no shift in aggregate demand in the flexible price specification, there was no pressure on prices or wages to begin with. Since aggregate demand is unchanged, adding price and wage rigidity has no effect on the outcome.

III.E. International Labor Mobility

To proxy for labor mobility, we adopt a high Frisch elasticity of 10 (relative to the baseline elasticity of 0.5) and keep all other parameters set at the values in the baseline case (see the fourth panel of table 2). With labor mobility, changes in the real after-tax wage produce much larger changes in labor supply. In effect, because some workers can leave Greece and work abroad, and we assume that emigrating workers do not remit their labor earnings, domestic income (gross national product) falls one-for-one with the reduction in labor income. Notice that there is no contradiction with the microeconomic estimates of relatively low Frisch labor supply elasticities in our specification. The workers could all have Frisch elasticities of zero but still be willing to move abroad to avoid enduring the domestic policies that Greece adopts. Relative to the baseline model, an economy with mobile labor requires much greater policy changes to meet the revenue target. To afford a flow payment to foreign creditors of 1 percent of GDP, government spending has to fall by 1.3 percent and output has to fall by 1.5 percent in the short run. In the long run, output remains below trend by roughly 0.81 percent. Tax policies are similarly contractionary; the labor tax policy causes a long-run reduction in output of 4.06 percent, while the capital tax approach causes a long-run reduction of more than 9 percent.

IV. Discussion and Caveats

The policy options considered above were each structured to be sufficient to increase the primary balance by 1 percent of GDP. Greece's actual obligations are closer to 4 percent of GDP on a flow basis. Given the linearity of our model, we can get a rough sense of one policy combination that would come close to fully "resolving" Greece's debt burden by simply summing across the rows of the four policy options. For instance, in the baseline model, a decrease in government spending of 1.21 percent of GDP together with a 2.2 percentage point increase in the consumption tax,

a 2 percentage point increase in the labor tax, and a 7 percentage point increase in the capital tax would raise enough revenue to afford a 4 percent flow payment to Greece's creditors. This would entail a short-run decrease in GDP of approximately 9 percent and a long-run (permanent) decrease of 5 percent. If labor were mobile, the decrease in long-run GDP would be *nearly 15 percent.*

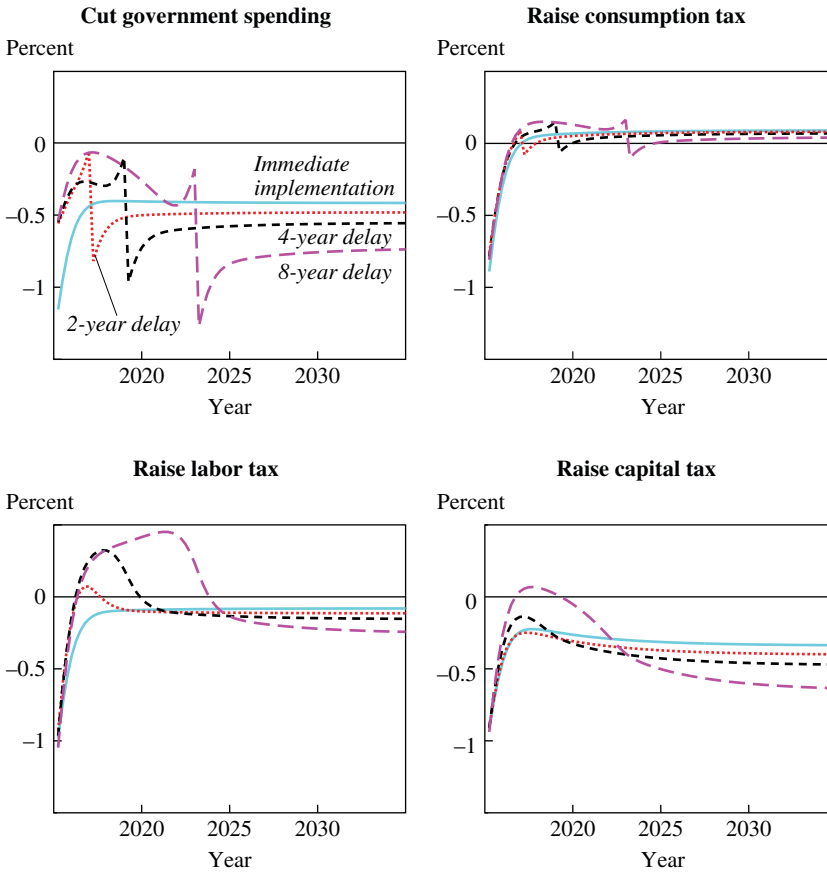
There are several important real-world issues that we have suppressed in our analysis. First is the credibility of the Greek government's carrying out future reforms. The policy simulations above illustrate the severity of the adjustments that Greece is facing, and it simply might be too much to expect that Greece will have the political willpower to follow through with such policy changes. Second is the issue of tax compliance in Greece. The tax changes we outlined above were changes in the effective tax rates, not increases in the statutory rates. To the extent that Greek firms and workers can avoid statutory tax increases, the required increase in the statutory rates will need to be even greater than the rate increases we analyzed. Finally, in the simulations we considered, Greek workers and firms are surprised by the revelation of the extent of government debt at the same instant that the Greek government sets out on a new policy path. In reality, labor and investment adjustments are already under way in Greece, influenced by the expectations of the policy choices that Greece might make.

V. Costs and Benefits of Delaying Austerity

Given the current weakness of the Greek economy, a natural question is whether the necessary fiscal adjustments should be delayed. Delay would allow prices and wages to adjust in anticipation of the policy changes, hopefully mitigating the disruptive effects of sluggish nominal adjustments. On the other hand, there is a cost to delay, because any interim deficits before the austerity policies are implemented must be financed. In the analysis up to this point, we have implicitly assumed that there is no risk of default and that temporary shortfalls in revenue relative to the target could be financed at the eurozone (real) interest rate of 2 percent. The interest rate in the event of a delay would likely include a risk premium. Here we assume that incremental debt due to delay carries an interest rate of 6 percent (a risk premium of 4 percent over the 2 percent baseline rate).

To illustrate the trade-offs that Greece faces, we use the model to simulate four delay scenarios for the four policy adjustment options discussed above. For each policy adjustment (cutting government spending,

Figure 2. Effects of Policies on GDP under Different Delay Scenarios[a]



a. Plots simulated GDP for different polices and different delay horizons. Each panel considers a different policy option, and each line in a given panel considers a different delay horizon.

raising consumption taxes, raising labor taxes, and raising capital taxes), we simulate the response to a policy that goes into effect immediately and compare it with responses to policies with a 2-year delay, a 4-year delay, and an 8-year delay. The 4-year delay is similar to the actual recommendation in the Memorandum of Understanding, which called for increases in the primary balance of –0.25 percent in 2015, 0.5 percent in 2016, 1.75 percent in 2017, and 3.5 percent thereafter (European Commission 2015, p. 6). Figure 2 shows the simulated trajectories for

Greek GDP under each of these scenarios. The top left panel shows the responses to cutting government spending.

In all cases, there are clear short-run benefits to delay. Looking at the 2015–20 period, the delayed policies feature GDP that is roughly 0.5 percent greater than policies that go into effect immediately. The exception seems to be the consumption tax, for which the benefits of delay are small in all cases. Unfortunately, there are also clear long-run costs to delay. In every case, output in the long run is lower than otherwise by as much as 0.25 percent. How the Greek government would weigh these costs and benefits is unclear. It is also unclear what Greece's options for delay really are. If Greece can roll over its debt at a low interest rate, then the case for delay becomes stronger. Indeed, if it can roll over debt at below-market rates, this would be a form of debt forgiveness.

VI. Other Considerations

Here we consider three additional factors that may influence the conclusions. Specifically, we analyze the role of long-run economic growth, the possible benefits of structural reforms, and the potential benefits of debt write-downs and sales of assets owned by the Greek government.

VI.A. Economic Growth

Our analysis so far has adopted a pessimistic view, namely that Greece will not grow at all in the coming decades. A zero percent growth rate is consistent with Greece's recent experience (since 2000, real growth has actually been negative) and also consistent with some long-term forecasts (McQuinn and Whelan 2015). However, if Greece *does* grow over time, this would allow the government to raise more revenue for any given change in tax rates. To a rough approximation, if the annual discount rate for Greek debt is r and the long-run annual growth rate is g, then raising an amount equivalent to a perpetual payment of roughly 1 percent of Greece's 2014 GDP would require an adjustment that is only $(r - g)/r$ times as large as the adjustments reported in table 2. Thus, if the delay rate were indeed 6 percent, as we considered in our analysis of the delay scenarios above, and the growth rate were roughly 2 percent, then the tax changes (and spending cuts) would need be only two-thirds as large as those considered in table 2. The International Monetary Fund (2015, p. 9) reports that its revised growth assessment for Greece going forward is approximately 1.5 percent, though it still describes this projection as "ambitious."

VI.B. Structural Reforms

Part of the August 2015 negotiations included specific requirements that Greece undertake a variety of structural reforms to modernize and liberalize labor markets, reform its pension system, reduce government regulation, and promote competition in product markets (European Commission 2015). Our analysis assumed that these structural reforms either would not be undertaken or would yield only a limited improvement in the functioning of the Greek economy. On the other hand, if the proposed structural reforms do result in substantial improvements to economic conditions, this would potentially reduce the fiscal pressure on the Greek economy. Quantifying the anticipated payoffs associated with the proposed structural reforms is extremely difficult. One recent attempt at doing so is that of Kieran McQuinn and Karl Whelan (2015), who use a neoclassical growth model similar to ours together with some plausible assumptions on the effects of major structural reforms to quantify the potential effects of the reforms. Most of the policies considered by McQuinn and Whelan (2015) focus on increasing labor supply, either by reducing labor market regulation or by reducing the generosity of state pension systems. Under the most optimistic scenarios, these reforms could push Greek labor market outcomes toward the labor market performance of the most productive countries in Europe. Were such an improvement to occur, it would raise Greek GDP substantially in the long run and raise long-run revenue.

VI.C. Debt Write-Downs and Asset Sales

One final consideration is the reduction of the debt burden by either debt forgiveness or through the sale of Greek assets. Both of these options would work directly to reduce the overall debt burden. Initially, the International Monetary Fund (2015) hoped that Greece could raise as much as €23 billion through the sale of various state holdings. While this sum is less than 10 percent of the overall total amount of debt payments, its effect on the present value of the debt would be substantial. For example, if we discounted future Greek debt payments at a 4 percent annual rate, asset sales of €23 billion would allow Greece to reduce its annual debt payment from 4.1 percent of GDP to roughly 3.5 percent of GDP.

VII. Conclusions

This paper provides a number of estimates of the impact of alternative fiscal adjustments that would enable Greece to increase its primary balance on a permanent basis by 1 percent of 2014 GDP. Under reasonable

assumptions, we show that (i) the required adjustments are very large and very painful, (ii) they are even larger when one takes into account realistic elasticities of the tax base, and (iii) they are larger still when one takes into account that Greece is a small open economy. There could be some short-term benefits from delaying fiscal adjustment, but delay would come at a relatively high price unless Greece's creditors were willing to provide additional finance at a relatively low interest rate. All of these adjustments become less painful under a scenario in which the Greek economy returns to a positive growth path. Whether proposed structural reforms can actually produce sustained growth remains to be seen.

References

Barattieri, Alessandro, Susanto Basu, and Peter Gottschalk. 2014. "Some Evidence on the Importance of Sticky Wages." *American Economic Journal: Macroeconomics* 6, no. 1: 70–101.

Blundell, Richard, and Thomas MaCurdy. 1999. "Labor Supply: A Review of Alternative Approaches." In *Handbook of Labor Economics* 3A, edited by Orley C. Ashenfelter and David Card. Amsterdam: North-Holland.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2011. "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins." *American Economic Review* 101, no. 3: 471–75.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113, no. 1: 1–45.

European Commission. 2015. "Memorandum of Understanding between the European Commission Acting on behalf of the European Stability Mechanism and the Hellenic Republic and the Bank of Greece." Brussels. http://ec.europa.eu/economy_finance/assistance_eu_ms/greek_loan_facility/pdf/01_mou_20150811_en.pdf

Heathcote, Jonathan, and Fabrizio Perri. 2002. "Financial Autarky and International Real Business Cycles." *Journal of Monetary Economics* 49, no. 3: 601–27.

House, Christopher L., Christian Proebsting, and Linda Tesar. 2015. "Austerity in the Aftermath of the Great Recession." Working Paper. http://www-personal.umich.edu/~ltesar/pdf/austerity.pdf

International Monetary Fund. 2015. "Greece: Debt Sustainability Analysis (Preliminary Draft)." IMF Country Report no. 15/165. Washington.

Klenow, Peter J., and Oleksiy Kryvtsov. 2008. "State-Dependent or Time-Dependent Pricing: Does It Matter for Recent U.S. Inflation?" *Quarterly Journal of Economics* 123, no. 3: 863–904.

Mankiw, N. Gregory, and Matthew Weinzierl. 2006. "Dynamic Scoring: A Back-of-the-Envelope Guide." *Journal of Public Economics* 90, nos. 8–9: 1415–33.

McQuinn, Kieran, and Karl Whelan. 2015. "Europe's Long-Term Growth Prospects: With and Without Structural Reforms." Working Paper no. 15/08. Dublin: Centre for Economic Research, University College Dublin.

Mendoza, Enrique G., Assaf Razin, and Linda L. Tesar. 1994. "Effective Tax Rates in Macroeconomics: Cross-Country Estimates of Tax Rates on Factor Incomes and Consumption." *Journal of Monetary Economics* 34, no. 3: 297–323.

Mendoza, Enrique G., Linda L. Tesar, and Jing Zhang. 2014. "Saving Europe? The Unpleasant Arithmetic of Fiscal Austerity in Integrated Economies." Working Paper no. 20200. Cambridge, Mass.: National Bureau of Economic Research.

Nakamura, Emi, and Jón Steinsson. 2008. "Five Facts about Prices: A Reevaulation of Menu Cost Models." *Quarterly Journal of Economics* 123, no. 4: 1415–64.

Trabandt, Mathias, and Harald Uhlig. 2011. "The Laffer Curve Revisited." *Journal of Monetary Economics* 58, no. 4: 305–27.

YANNIS M. IOANNIDES
Tufts University

CHRISTOPHER A. PISSARIDES
London School of Economics

Is the Greek Crisis One of Supply or Demand?

ABSTRACT Greece's "supply" problems have been present since its accession to the European Union in 1981; the "demand" problems caused by austerity and wage cuts have compounded the structural problems. This paper discusses the severity of the demand contraction, examines product market reforms, many of which have not been implemented, and their potential impact on competitiveness and the economy, and labor market reforms, many of which have been implemented but due to their timing have contributed to the collapse of demand. The paper argues in favor of eurozone-wide policies that would help Greece recover and of linking reforms with debt relief.

Greece joined the European Union (EU) in 1981 largely on political grounds to protect democracy after the malfunctioning political regimes that followed the civil war in 1949 and the disastrous military dictatorship of the years 1967–74. Not much attention was paid to the economy and its ability to withstand competition from economically more advanced European nations. A similar blind eye was turned to the economy when the country applied for membership in the euro area in 1999, becoming a full member in 2001.

It is now blatantly obvious that the country was not in a position to compete and prosper in the European Union's single market or in the euro area. A myriad of restrictions on free trade had been introduced piecemeal after 1949, with the pretext of protecting those who fought for democracy. These restrictions do not allow Greek companies to develop, adopt new technology, and grow into world-leading exporters—with the exception of shipping, which is subject to different rules because of its international nature. Professions are protected, and there is political interference with the

economy from the basic level all the way to the top, through state-controlled enterprises, rules and regulations that have accumulated over the years, and party-political appointments of officials who control licensing offices down to the local level. As a result, Greece is the most protected and monopoly-ridden economy in the euro area, and no attempt was ever made to reform the economy to raise its productivity to the level of its European partners.

These features of the economy should have been obvious to those examining the Greek case for entry into the European Union and the euro area. Whether they were obvious or not, however, is immaterial at the current juncture: The key point today is that they were ignored until the debt crisis of 2010. Entry into the EU kept Greece going through transfers, and entry into the euro area gave the country access to cheap finance, which provided funds for consumption and residential investment. This growth model was clearly unsustainable, but that fact was not exposed until after the onset of the global financial crisis and highlighted in a series of reports by international organizations, most of them associated with the troika's periodic reviews.[1]

It is clear that for the long-run viability of Greece's economy and survival in the eurozone the urgent need is for structural reform. But implementing deep and effective structural reform in an economy used to protectionism and political meddling meets with resistance at every level, leading to public protest, political instability, frequent elections, and the rise of political extremism. So, although since the onset of the crisis in 2010 several rounds of legislation went successfully through Parliament, the implementation of reform has been very poor. In practice there is no such thing as an independent public sector that will implement the reforms impartially according to any new legislation. In private conversations, economists brought in to advise the government on reform acknowledge that once they are in office, huge pressures are brought to bear on them to make exceptions that offset the impact of legislation to the point of complete irrelevance.[2]

The problem that should be occupying Greece's lenders is how to give incentives to achieve effective structural reform. Instead, their focus has

1. The *troika* refers to the European Central Bank, the European Commission, and the International Monetary Fund, which jointly administered the Economic Adjustment Programme for Greece (European Commission 2010). They have issued periodic reports, the latest of which is the European Commission's fourth review (European Commission 2014) and the International Monetary Fund's fifth review (IMF 2014). The European Commission's fifth review was interrupted in December 2014. See Hardouvelis (2015).

2. This is common knowledge in Greece. Most recently it was reiterated to us by Gikas Hardouvelis, who served as finance minister before the Syriza election victory in January 2015.

been on fiscal austerity, ever higher taxation, cuts in earnings, and debt sustainability, which has provided disincentives for reform. Reform is politically easier to implement and economically more effective when demand in the economy is at a healthy level, demand in the country's trading partners is expanding, the country's financial sector is in a position to support new ventures, and its fiscal authorities are in a position to help with infrastructure investments and private-public partnerships. All of these conditions characterized Germany when it embarked on its reform program in 2003–05, which was controversial at the time and included running a bigger budget deficit than allowed by the Maastricht Treaty in order to facilitate the transition to a new economic order. But these conditions have been denied to Greece, reinforcing its reluctance to implement reform and rendering ineffective the small number of successful reforms that have taken place, such as Enterprise Greece, which enables the speedy establishment of new companies for which currently there is no demand.

Our message is that ignoring Greece's problems when it first applied for membership in the single market and the euro area was a mistake and that the fiscal austerity and wage cuts that international lenders enforced upon Greece were also a mistake, one that has compounded the first mistake, because they stunt the necessary structural reform efforts. The distorted structure of the Greek economy, especially its concentration of economic activity in a small number of hands, introduces price inflexibilities, so the fiscal austerity and wage cuts have caused a catastrophic fall in demand. Such features are behind the International Monetary Fund's (IMF's) underestimation of the fiscal multipliers; the fall in demand is behind the deep recession and the rise in unemployment. It is also partly behind the failure of the 2012 sovereign debt restructuring to deal once and for all with the Greek debt.[3] However, the high debt is a *consequence*, rather than a *cause*, of the current situation. Although more debt relief now would enable Greece to implement reforms with more flexibility because it would somewhat relax the austerity, we believe that even a complete write-off would not transform Greece into a modern growing economy that could prosper in the euro area.

Greece's problem is one of both supply and demand. The supply problems have been present since the country's acceptance into the EU in 1981; the demand problems caused by austerity and wage cuts have added to the supply problems, making them worse.

3. See Zettelmeyer, Trebesch, and Gulati (2013) for the full details of the 2012 restructuring, which is also known as the 2012 PSI (private sector involvement).

The remainder of this paper is organized as follows. In section I we discuss the severity of the demand contraction and its impact on economic activity, which, as we argue, is largely due to Greece's distorted economy. In section II we discuss product market reforms and their potential impact on competitiveness and the economy. In section III we turn to labor market reforms, many of which have been implemented, and discuss their impact. In section IV we argue in favor of eurozone-wide policies that would help propel Greece on the road to recovery. We also discuss how linking reforms with debt relief could increase the motivation for reform. Section V concludes.

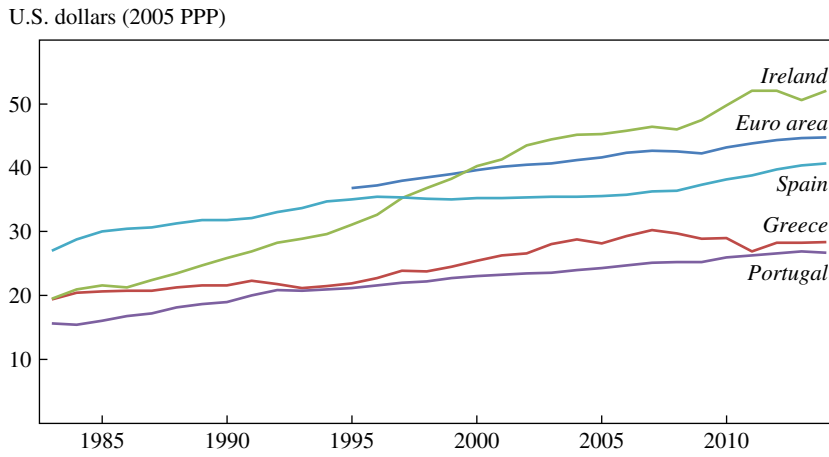
I. Demand Contraction

The main ingredients of the reform program forced upon Greece by its international lenders, as well as the thinking behind them, consists of three pillars: fiscal contraction to reduce the massive budget deficit and eventually pay off the debt; reductions in wages, pensions, and other costs to increase the competitiveness of Greek industry; and a structural reform program to modernize the economy and increase productivity.[4] The expectation of the institutional lenders was that the fiscal contraction would have very small negative multipliers, the "internal depreciation" that would result from reductions in unit labor costs would increase exports and also help domestic demand, and the structural reform would increase productivity and improve expectations about future prospects, giving access to more and cheaper finance for investment and output growth. In practice, fiscal austerity led to bigger negative multipliers than estimated by the IMF (as acknowledged in IMF 2013) and a bigger fall in output than expected. The internal depreciation helped exports to some extent, but the wage reductions that brought it about, combined with sticky prices, brought much bigger reductions in domestic demand (Pissarides 2013). And structural reform has been ineffective either because of its limited scale or because several of its dimensions have not been implemented.

Greece's large trade deficit has declined since the onset of the crisis, from 11.2 percent of GDP in 2009 to 2.3 percent in 2014. This was partly

4. An antiquated pension system was at the root of the fiscal explosion, requiring ever increasing contributions by the government. An ambitious reform effort, which recognized the adverse demographic trends and was initiated in 2001, was not completed. The 2010 Economic Adjustment Programme mandated extensive pension reforms. Such reforms continue to be key thorny issues in the negotiations by Greece with its lenders as of the time of writing.

Figure 1. Hourly Productivity, Euro Area, Greece, Ireland, Portugal, and Spain, 1983–2014



Source: OECD (http://www.oecd-ilibrary.org/statistics).

due to the absolute fall in imports (although they increased as a share of GDP), but also to an increase in exports from 19 to 33 percent of GDP (with exports of goods increasing from 8.4 to 17.3 percent of GDP). Hourly productivity declined during this period, in contrast to the rest of the euro area, but real average earnings fell by even more (figure 1 and table 1). As a consequence, real unit labor costs fell, improving the competitiveness of Greek exports despite the falling productivity.

Both the fall in wage costs and the fall in productivity were due to the collapse of aggregate demand and investment. Whereas fixed capital formation in the euro area in 2014 was at about 19 percent of GDP, only slightly below the precrisis levels, in Greece it collapsed from more than 20 percent in the precrisis years to 16.3 percent in 2009 and to 8 percent in 2014, with residential construction accounting for a large portion of the collapse.

According to the Organization for Economic Cooperation and Development's (OECD's) 2013 economic survey of Greece (OECD 2013, p. 53), the adjustment program has so far failed to restore price competitiveness, growth, and public debt sustainability, and the fiscal contraction has deepened the depression. There is no doubt that the troika has made errors in policy design. The fiscal multiplier assumed in the design of the program was much smaller than might have been at work, with a value of 0.5 instead of some value above 1, which is more widely used now. This would have

Table 1. Real Average Earnings, Greece, Ireland, Italy, Portugal, and Spain, 2009–14[a]

Year	*Greece*	*Spain*	*Ireland*	*Italy*	*Portugal*
2009	100	100	100	100	100
2010	93	98	99	101	100
2011	88	96	98	99	97
2012	85	93	97	96	93
2013	80	94	97	96	95
2014	81	93	99	97	93

Source: OECD (http://www.oecd-ilibrary.org/statistics).
a. Values indexed to 2009 = 100.

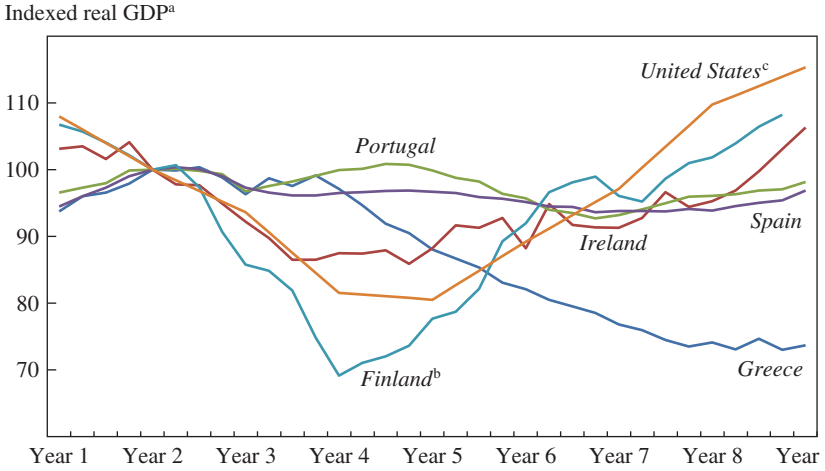
had a large impact on the estimated contraction, with everything else constant.[5] Of more importance, however, are the peculiarities of the Greek economy, which had largely been overlooked. The Greek economy is subject to more frictions and is less open than other euro area economies in crisis, such as Ireland and Portugal.[6] In such an economy, prices do not fall, and the decrease in aggregate demand brought about by wage decreases translates into a contraction of aggregate activity and unemployment (Pissarides 2013). Wage reductions in Greece were reflected in greater increases in profit margins rather than reductions in prices.

As a result of the demand contraction and the large multipliers, output declined in Greece dramatically more than in other countries under stabilization programs (figures 2 and 3). In early 2013, GDP was 25 percent below its 2008 level, in sharp contrast to the drop of 10 percent or less in the other countries. Figure 2 also compares Greece with Finland during

5. There are two different positions taken by IMF staff. Bi, Qu, and Roaf (2013, p. 26) argue that projections would not have been very different if higher multipliers had been assumed. Instead, they claim that the error was in the forecast of potential GDP, which anticipated a contraction of about 7 percent instead of the observed 20 percent. Olivier Blanchard attributes the underestimation to substantial evidence that turned out to be misleading in an environment of near-zero interest rates and monetary policy that cannot offset the effects of the fiscal stance (IMF 2015a). IMF (2010, p. 93) examines in depth the different aspects of macroeconomic effects of fiscal consolidation, finding that "fiscal consolidation typically reduces output and raises unemployment in the short term." The process is the more painful the less monetary stimulation is available and the greater the reliance on taxes, and both of those conditions have been features of the Greek stabilization.

6. Average total exports (goods and services) as a percent of GDP were as follows for 2008–12: Greece, 23.4; Ireland, 93.1; and Portugal, 33.2. Greece's intra-EU exports and imports, as shares of respective totals, are the lowest among comparably sized EU countries. The share of intra-industry trade is probably not very large, thus the greater the need for adjustment through the exchange rate and the smaller the productivity-like potential benefit from "task trading" (Grossman and Rossi-Hansberg 2008).

Figure 2. Real GDP, Greece, Ireland, Portugal, and Spain, 2007–15, and during Depressions in Finland and the United States



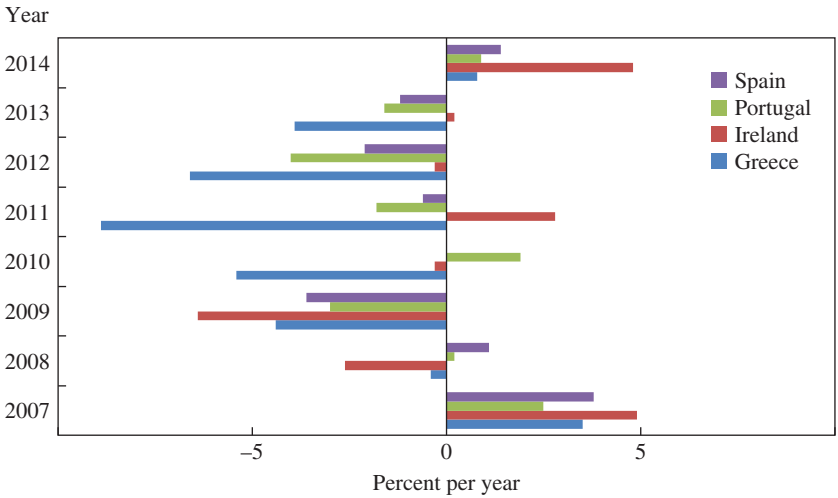
Sources: Eurostat (http://ec.europa.eu/eurostat/data/database), U.S. Bureau of Economic Analysis.

a. For Greece, Ireland, Portugal, and Spain, Year 1 = 2007 and data are indexed to 2008 = 100; for Finland, year 1 = 1990 and data are indexed to 1991 = 100; for the United States, year 1 = 1929, and data are indexed to 1930 = 100.

b. Depression in Finland spanned 1990–97; data plotted quarterly are 1990Q1–1997Q4.

c. Depression in the United States spanned 1929–38; data plotted annually are 1929–37.

Figure 3. Real GDP Growth Rate, Greece, Ireland, Portugal and Spain, 2007–14



Source: Eurostat (http://ec.europa.eu/eurostat/data/database).

its "great depression" of 1990–96 and the United States during the Great Depression of 1929–38.

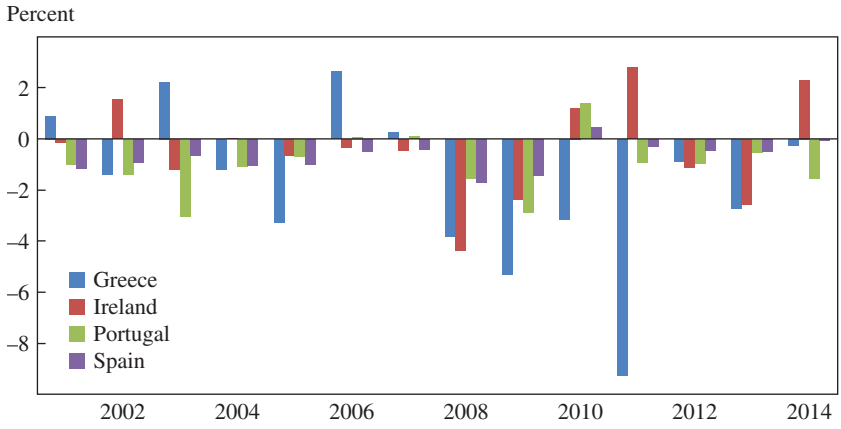
Even more striking is Greece's performance in relation to other European countries. Between 2009 and 2012, Greek income per capita fell from 96 to 75 percent of the EU average, below its level in 1995. Between 2010 and 2012, annual real income contraction in Greece was almost 3 percentage points deeper than in the OECD and troika forecasts, while nominal GDP contraction was 3.5 percentage points deeper. Relative to the 3.8 percent trend growth rate in the period 1997–2007, the fall of GDP in 2007–13 was 38 percent (Müller, Storesletten, and Zilibotti 2015). Much of the gain made since Greece's eurozone accession has been lost. From 2009 to 2014, real GDP per person declined by 21 percent; relative to the EU-28 average, it declined from 93 percent in 2003 to 70 percent in 2014.

II. Structural Reforms and Competitiveness

Given the sticky prices and barriers to entry, instead of reversing the recession, the fall in wages has so far contributed to it by reducing aggregate demand, as has the fall in unit labor costs by failing to spur competition. In such circumstances it makes much more sense to target product market reforms, which could improve price flexibility and the structural competitiveness of the Greek economy. Labor market reforms are also essential, but they could come later, once the economy was performing well, since they would be easier to implement. Labor market reforms are resisted by workers and their unions and they can be disruptive, both politically and economically. Product market reforms are resisted by the professions and by owners of capital, who ultimately are more likely to comply if sufficient compensation to the losers is given.

The issue of the urgent need for structural reforms has figured prominently in all discussions of the Greek crisis. Product market deregulation can affect growth in two important senses that are not typically clarified in public debates. One effect is from abolishing monopolistic and monopsonistic structures and eliminating barriers to entry, which would benefit productivity. A second effect is from market deregulation's impact on total factor productivity (TFP) growth. Boosting TFP growth is essential if Greece is to recover and catch up with other European countries, but since the country's entry into the reform program, TFP growth has declined in Greece by much more than in other program countries (figure 4). The fundamentals are against the revival of TFP growth, mainly because of too little spending on education and on research and development, as well

Figure 4. Total Factor Productivity, Greece, Ireland, Portugal, and Spain, 2001–14



Source: The Conference Board.

as poor connections between universities and industry (Pissarides 2015). Greece's spending on research and development increased slightly between 2002 and 2013, from 0.6 percent to 0.8 percent of GDP, but these numbers are well below the figures for the average of the euro area during the same period, rising from 1.8 to 2.0 percent.

The first channel through which deregulation can affect growth has served as a key objective of deregulation efforts that break barriers to competition, and it has been an important part of the reform programs in the European Union. Its main impact is a jump in potential output, moving the country closer to the technological frontier. The second channel is an important component of the EU's Europe 2020 growth strategy, which includes a Digital Agenda for Europe. The impact of this reform is mainly on growth through the beneficial effect of structural reform on research and development and on trade competitiveness, and is thus vastly more important because it is long lasting.

The importance of structural reforms is enhanced in the presence of downward nominal wage rigidities, fixed exchange rates, and high debt levels, which all characterize the members of the euro area. For this reason, they have been repeatedly emphasized during the European crisis by politicians and leading figures in European institutions. Notably, as Mario Draghi (2015), president of the European Central Bank (ECB), eloquently argues, slow-adjusting countries within the eurozone are likely to suffer higher unemployment, which can become entrenched and structural, whereas

structural reform can bring the European economies closer together and thus improve the chances of success of a uniform monetary policy.

Writing in 2009, an IMF team identified many key weaknesses in the Greek economy (Moreno-Badia, Traa, and Velculescu 2009). They argued that the imputed equilibrium real exchange rate was overvalued relative to fundamentals and implied a competitiveness gap of 20 to 30 percent; that the weaknesses of Greek labor markets were glaring, with low employment rates, especially for females and the young; that relatively high employment protection legislation and structural impediments, including cumbersome business practices and high costs to start a business, hindered product market performance; and that internal competition was insufficient due to high regulation and limited liberalization of utilities, which implied higher energy costs and poorer supply chains for the whole economy.

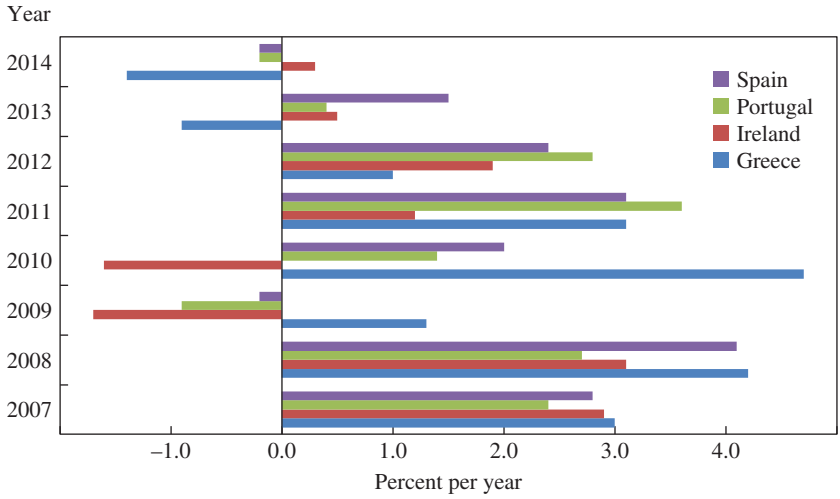
Those observations motivated an important part of the Economic Adjustment Programme for Greece, which was agreed to in May 2010 (European Commission 2010). Drawing examples from other countries that had undergone reform, the IMF team, referred to above, concluded that were Greece to move toward best practices in each of those areas, the corresponding employment gains could be significant, estimated between 5 and 10 percentage points.

As the 2010 Economic Adjustment Programme gave way to a second program in 2012, the need for reforms was further specified and the troika's demands on Greece became more pressing. But product market reforms were not given priority and whatever interventions were implemented, they amounted only to partial dealing with some of the problems. Price inflation had increased from 1.3 percent in 2009 to 5 percent in 2010, before going down to −1 percent in 2013 (figure 5), while wage growth decreased sharply from 2009 to 2013 (figure 6). The net effect on the real wage has been dramatic, as table 1 and figure 7 show, with real wages falling by more than a quarter since 2007. The labor share fell from 56 percent in 2009 to 48 percent in 2013.[7] At the same time, reformers had to tackle fierce opposition from vested interests and were hampered by a lack of political commitment. Product markets continue to be dominated by oligopolies, with numerous barriers to entry protecting incumbents.

As of the time of this writing, many of these conditions remain unsatisfied and are listed as part of the latest agreements as "prior actions" (European

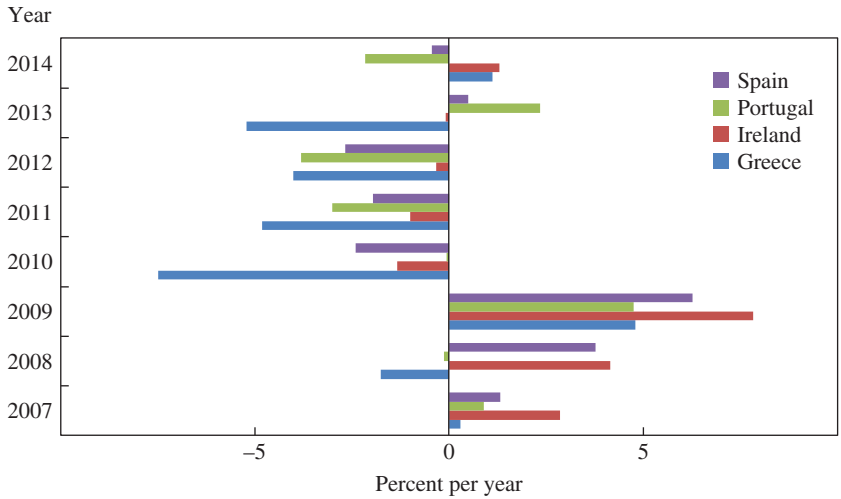
7. As reported in the 2015 AMECO Database (http://ec.europa.eu/economy_finance/db_indicators/ameco/index_en.htm).

Figure 5. Inflation HICP,[a] Greece, Ireland, Portugal, and Spain, 2007–14



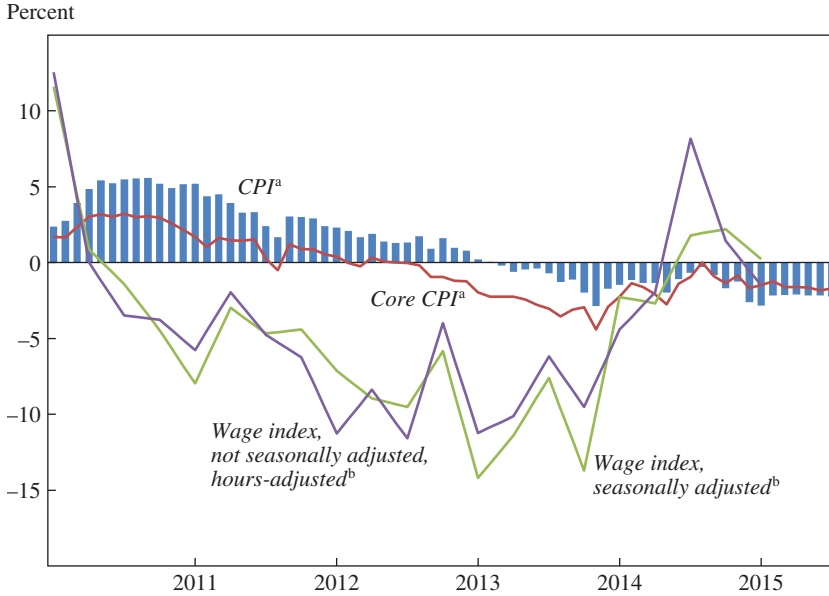
Source: Eurostat (http://ec.europa.eu/eurostat/data/database).
a. Harmonized Index of Consumer Prices.

Figure 6. Wage Growth Rate, Greece, Ireland, Portugal, and Spain, 2007–14



Source: OECD (http://www.oecd-ilibrary.org/statistics).

Figure 7. CPI and Wage Trends, Year-over-Year, 2010–15



Source: ELSTAT (http://www.statistics.gr/en/statistics/eco).
a. Consumer price indexes are measured monthly.
b. Wage indexes are measured quarterly.

Commission 2015). A slow improvement did take place in the period leading up to the January 25, 2015, parliamentary elections. Some international indicators have improved, including Greece's ranking in the Global Competitiveness Index, from 96 in 2012 to 81 in 2014.[8] Similarly, Greece improved in the World Bank's ease of doing business ranking, moving from 72 in 2014 to 61 in 2015,[9] and in the OECD's Product Market Regulation Index it improved from 2.21 in 2008 to 1.74 in 2013.[10] But such indexes are still a long way behind the levels that one would consider to be suitable for an economy on a convergence path with the rest of Europe.

Regulation remains high and is an obstacle to catching up with the rest of the euro area. Using the IMF's Global Integrated Monetary and Fiscal (GIMF) model, Stephanie Eble and others (2013) calculate that policies

8. See table 3 in Schwab (2012) and Schwab (2014).
9. See table 1.1 in World Bank (2013) and World Bank (2014).
10. Index scale is from 0 to 6, with 0 being least restrictive and 6 being most restrictive. See table 2 in Isabell Koske and others (2015).

that could close roughly half the gap between Greece and the rest of the eurozone in measures of product and labor market regulation would raise real GDP by about 4 percent after 5 years, rising to 10 percent in steady state. Janos Vargas, Werner Roeger, and Jan in 't Veld (2013) show that reforms in product and labor markets and improvements in knowledge and innovation (as measured by research and development subsidies and the skill composition of the labor force) account for about 78 percent of the gap between Greece and the top euro area performers in 2012, measured in percent deviation from the baseline. Reforming the product market yields the highest GDP gains in Greece.[11]

Dimitris Papageorgiou and Evangelia Vourvachaki (2015) use calibration techniques with Greek data to measure the macroeconomic and growth impacts of structural reforms. They find strong positive long-run gains from implementing structural reforms in the product and labor markets: 10-percentage-point reductions in nontradable prices and private sector wages lead to 9 percent gains in GDP. The results also suggest that the mix of fiscal policies helps determine the impact of simultaneously implemented structural reforms and debt consolidations in the short to medium terms. In the long run, the gains of the two policies complement each other, with additional GDP gains being in the range of 0.4 to 4 percent. By improving the economy's permanent productive capacity, structural reforms also improve the tax base. Although reforms create additional fiscal space, thus conferring benefits in the long run, they generate losses in the short and medium run, so it is not only pace that matters but also the mix of fiscal tools.

Ultimately, a bigger impact of structural reform in Greece would be one that worked its way through higher TFP. With lower entry barriers and less state control, manufacturing industries could catch up with best-practice technologies faster. Giuseppe Nicoletti and Stefano Scarpetta (2003) find that changing governance structures—for example, through privatization—would bring along improved competitive pressures and

11. That structural reforms are associated with growth may be easily verified by means of Barro regressions with the growth rate of per capita income as the dependent variable and initial income and a whole host of variables measuring structural reforms as independent variables. Cheptea and Velculescu (2014) report that 121 such structural reform variables, which include measures of corruption, research and development, corporate governance, and infrastructure, are statistically significant. Such regressions are known not to be causal, but they are nonetheless qualitatively informative about the scope for improvement in growth performance from narrowing the gap between actual and "benchmark" performance. For example, these authors report that the "average growth effect" from institutional reforms is at 1.3 percent for Greece, the third highest effect in the EU, after Bulgaria and Romania, both at around 1.5 percent.

entrepreneurial incentives. This would especially be the case if competition were promoted in the markets where privatized industries operate, such as energy, telecommunications, and transport, since companies in these industries provide inputs to the entire economy. In particular, they estimate that a gradual move (over 10 years) to the OECD-wide average share of state-owned firms in total value added would boost annual TFP productivity growth by about 0.7 percentage point in Greece. They also estimate that entry liberalization in service industries would boost annual TFP growth in the overall business sector by about 0.1–0.2 percentage point in countries like Portugal, Greece, and Italy.

Against the background during the 1975–2003 period of some OECD countries showing impressive TFP growth performance and others showing the opposite, Christopher Kent and John Simon (2007) find that changes in TFP growth are positively correlated with information and communications technology (ICT) spending as a share of GDP. They also find that the share of ICT spending is negatively correlated with the level of product market regulation. Greek industry is dominated by micro-firms, which account for about 58 percent of the nonfinancial business community, while larger firms account for only 13 percent. It therefore does especially poorly in ICT penetration.[12] The ongoing restructuring of the Greek economy (discussed below) is thus promising in this regard.

Underlying Greece's competitiveness problem is the fact that the Greek economy does not mobilize enough knowledge as expressed through the knowledge composition of the country's exports relative to those of the rest of the world. Among a sample of 128 countries, Ricardo Hausmann and others (2013) found that Greece had the largest gap between its level of income and the knowledge content of its exports. The same set of calculations suggest that Greece ranks second only to India in terms of how easy it would be to move to exporting more complex goods (Hausmann 2012). The average domestic value added of Greek exports over the period 1995–2008 was at around 10 percent of GDP, ranking it the lowest among all EU countries (and Turkey), which suggests huge room for

12. IMF (2015b), chapter 3, reports empirical results with OECD country data that show somewhat mixed effects of structural reforms on TFP, but generally the impact is positive. Econometric estimates suggest that lower product market regulation and more intense use of high-skilled labor and ICT capital inputs, as well as higher spending on research and development activities, contribute positively and with statistical significance to total factor productivity. The effects vary across sectors and are typically larger the closer the sector is to the technological frontier. For example, product market deregulation has larger positive total productivity effects in the services sector, but high-skilled labor and research and development expenditure have the strongest effects in ICT-related sectors.

improvement through vertical supply links' gaining a greater role (Rahman and Zhao 2014). Greece is the economic neighbor of some of the world's most advanced countries, a circumstance that ought to facilitate diffusion of innovations and technologies.

Despite all these potential gains from product market reforms, there is resistance to the adoption of any reform. We believe that if implementation is to succeed, the reforms ought to be "owned," beyond the sphere of politics, by the groups that will implement them.[13] For example, product market reforms that free up competition in trades such as taxi transportation and pharmacies, let alone more far-reaching ones that may be threatening to broader groups of the population, must be eased in gradually and give affected workers alternative means of support in the transition, since removal of entry barriers and legislated mark-ups will shrink the affected sectors. Indeed, the rationale of IMF support to restructuring countries is to provide a cushion to the losers and help them in the transition to a new economic order. But despite the availability of ample finance for this purpose and the passage of more than five years since the initial agreement with the troika (in May 2010) to free up competition, several professions continue to jealously guard their privileges by restricting access to licensing and only slowly letting go of gross overbilling practices for services provided through public sector projects.[14]

We should also note that even if reforms were successfully implemented, there would still be a time lag of about 3 to 4 years before they had an impact on the real economy. We know this from the experience of other countries that have reformed. In Germany, for example, the Hartz reforms were implemented over three phases, from 2003 to 2005, with full cooperation among unions, employers, and government, but their impact only showed up in the German labor market starting in 2007 (Pissarides 2013). Research by Christian Dustmann and others (2014) also emphasizes the importance of *trust* in wage-setting institutions and other aspects of employment decisions in driving productivity improvements. With the current absence of trust and resistance to reform, Greece will require help

13. The Memorandum of Understanding between the European Commission, the Greek government, and the Bank of Greece—which the Greek government recently passed through Parliament—specifically states: "Success requires ownership of the reform agenda programme by the Greek authorities" (European Commission 2015, p. 4).

14. The prevalence of overbilling is common knowledge in Greece, and it is one of the reasons businesses were keen to get public sector contracts—only to discover recently that payments are not forthcoming because of the state of public finances. Business people who asked not to be named told us that they are discounting public sector contracts by more than 50 percent and still making substantial profit.

from the international institutions for a longer period than the current three-year agreement signed on August 19, 2015 (European Commission 2015), even if the country effectively reformed starting now.

III. The Impact of Labor Market Reforms

Labor market reforms have been given greater priority in Greece than product market reforms, mistakenly in our view. Whether this was because successive Greek governments found it easier to reform labor markets than product markets or because the troika insisted on them is a moot point.[15] The sequencing of reforms brought about the large fall in wages ahead of any price adjustment, with the demand consequences that we have outlined.

Several changes have taken place in the structure of labor market institutions. The Greek labor market used to exist in a rigid framework characterized by numerous firing restrictions, restrictions on overtime work, and the minimum wage. Collective bargaining at the national level, the industry level, and the firm level determined many outcomes, and so did many other frictions. Relaxing all those restrictions does facilitate sectoral reallocation, but compliance with the various regulations has not been universal. Apart from some large firms, the predominant and most widespread firms in Greece are family-owned and very small, making it easy for them to circumvent labor laws. Labor inspections were minimal. Moreover, the shadow economy in Greece is much larger than in the rest of the eurozone, estimated to make up 24 percent of GDP as compared with the eurozone's 15 percent (European Commission 2013). These features of the Greek economy also explain why the reform of the collective bargaining law has had a limited impact so far, largely because it involved a small number of firms (Lyberaki, Meghir, and Nikolitsas forthcoming).

Of greatest macroeconomic significance is the sharp reduction of the minimum wage, the decentralization of wage bargaining to the firm level, and the extensive relaxation of employment protections. The minimum wage was reduced in both the public sector and the private sector by 22 percent in February 2012 from a monthly rate of €751 to €581. The aggregate contraction and the increased flexibility of labor markets have been associated with a large decline in unit labor costs, which have fallen

15. Gikas Hardouvelis informed us (in a private communication) that from his experience as chief economic adviser to Lucas Papademos as prime minister of Greece from November 2011 to May 2012, during negotiations in spring 2012 it was clear that the troika was insisting on labor market reforms to precede product market reforms, arguably because of convenience.

by about 20 percent since 2009.[16] The real exchange rate based on unit labor cost has depreciated by 16.5 percent since 2009, though the CPI-based rate has depreciated by only 5.6 percent since 2009.

Did the extensive labor market reforms undertaken in Greece have any positive impacts? The impacts were on the whole small and fragile. In 2014, the Greek labor market started to show signs of recovery. Employment contracted at a slower rate than in 2013, by –0.9 percent year-over-year in 2014Q1, as compared with –2.9 percent in 2013Q4 and –4.9 percent in fiscal year 2013. The evidence from business-cycle and forward-looking indicators signaled, in July 2014, an expansion in employment in 2014Q3, which occurred at 1.4 percent year-over-year in 2014Q3 and increased to 1.5 percent year-over-year in 2014Q4 (National Bank of Greece 2014a, 2014b, 2015).

More than two-thirds of employment losses in the private sector (730,000 jobs) had been due to the closure of about 220,000 small firms (30 percent of the existing small enterprise count) together with layoffs in that sector. Most of those jobs were lost in firms with a domestic orientation and with less flexible labor market structures, reflecting the Greek economy's adjustment to a greater role for larger and more export-oriented firms. Indeed, job losses in medium-sized and large firms were half those experienced in small firms (–17 percent cumulatively since 2008). The contribution to total turnover made by larger firms (those with turnover exceeding €50 million) increased from 27 percent of turnover in 2008 to 36 percent in 2013, bringing Greece closer to the EU average of 43 percent in 2013. And after several years of decline, the number of firms started growing in 2014, due both to existing firms getting back into business and to new firms being established (National Bank of Greece 2015).

The fact that wages fell by 23 percent in the period 2009–13 and employment contracted by 24 percent cumulatively during that period caused the wage share in the economy to fall to 48 percent of GDP, 13 percentage points below its 25-year average. On the other hand, capital income, mainly comprising the gross operating surplus of the business sector, has proved more resilient—reflecting an ongoing corporate restructuring and lower labor costs—declining by 19.7 percent in the 5 years leading up to 2013. These developments are increasingly favorable for new hiring, since improving business profit margins should lead to higher investment and business expansion. By linking employment to corporate profitability and output growth, the National Bank of Greece (2014b) forecast that it should

16. From the 2015 AMECO Database; see note 7.

be profitable for Greek firms to increase their employment by an average pace of 2.5 percent per year until 2020, or 19.6 percent cumulatively between 2014 and 2020, which would add up to 720,000 new employment positions during the period, pushing the unemployment rate below 21 percent in 2016 and 12 percent by the end of 2020; "Such employment creation will clearly depend on the timely implementation of the programme, including its growth-enhancing structural reform agenda" (National Bank of Greece 2014b, p. 1). However, these optimistic assessments were not realized, with the uncertainties introduced by the change of government in January 2015 playing a critical role.

Having examined the Greek case in some detail, it is natural to wonder whether or not the EU's institutional environment is more conducive than Greece's to the implementation of reforms. Alberto Alesina, Silvia Ardagna, and Vicenzo Galasso (2010) examine the linkages between the adoption and facilitation of structural reforms in the euro area, using both theoretical and empirical methods. They find that within the small sample of 11 countries that they work with, the euro has indeed been associated with an acceleration of product market reforms, which did come first, but not with labor market reforms.

IV. Dealing with Greece's Large Debt

A central tenet of the Greek adjustment program is that it would enable Greece to become sufficiently competitive so as to regain access to international capital markets. With its external trade balance being persistently negative, a central hope is that economic restructuring will alter fundamentals sufficiently to achieve an improved goods balance. The latter outcome, along with stronger earnings in Greece's traditional mainstay of tourism, will narrow the gap to be covered by capital flows in the form of fresh borrowing and investment. Therefore, a central question is to assess the main forces affecting Greece's accumulated external liabilities so as to be able to predict its ability to deal with its large debt.

The mainstream view is that Greece, Ireland, Italy, Portugal, and Spain have accumulated external liabilities due to their loss of competitiveness following relative increases in their unit labor costs. Ruo Chen, Gian Maria Milesi-Ferretti, and Thierry Tressel (2013) question this view by pointing to factors that are seemingly external to those countries and have affected them as a group. First, among European economies there has been an asymmetric trade interaction with emerging Europe, fast-growing China, and oil exporters. Germany has captured fast-growing markets

such as China for its exports and has integrated its production chains with central and eastern Europe, a factor that was also decisive in its ability to expand production without incurring domestic wage raises. Second, during 2000–09 the real exchange rate appreciation in those eurozone periphery countries reflected substantial nominal exchange rate appreciation.

If this is a correct diagnosis, a more accommodative ECB policy, by strengthening growth in the northern eurozone, would "lift all boats" in the eurozone periphery (Pissarides 2013). In addition, to the extent that the debtor countries are affected by the changing terms of trade in roughly the same manner, they could also benefit from eurozone-wide policies to further improve their competitiveness. That is, infrastructure investments and spending on research and development aimed at improved competitiveness will generate spillover effects, while such spending will benefit from larger multipliers than were typically assumed in the design of stabilization programs in the eurozone periphery.[17]

The findings of Chen, Milesi-Ferretti, and Tressel (2013) are not the entire story, however, at least as far as Greece and Spain are concerned. During the years 1999–2007, both countries experienced economic booms financed by borrowing and increasing exports (Galenianos 2015). In particular, Greek exports rose from 19 to 22 percent of GDP, increasing at one of the fastest rates of any eurozone country during that period. However, imports increased even faster, outrunning exports. The current account balance, as a share of GDP, had averaged nearly –8 percent for Greece during the period 1999–2008; in 2014 it was down to –2.6 percent. The Greek program can boast of success in the external trade area, as we have already argued, but despite the large fall in the trade deficit more needs to be done in view of the country's obligation to service its huge debt.

The ECB's Expanded Asset Purchase Programme, while principally aimed at offsetting deflationary pressures within the eurozone, can also improve competitiveness in the eurozone as a whole. That is, because of

17. There have been prominent voices, including that of Olli Rehn, European commissioner for economic and monetary affairs and the euro from 2010 to 2014, in favor of policy initiatives by the European North. As Rehn (2013) put it: "As the two largest eurozone economies, Germany and France together hold the key to a return to growth and employment in Europe. If Germany can take steps to lift domestic demand and investment, while France embraces reforms to its labour market, business environment and pension system to support competitiveness, they will together do a great service to the entire eurozone—providing stronger growth, creating more jobs and reducing social tensions." Most recently, Blanchard, Erceg, and Lindé (2015) have quantified substantial benefits for the periphery of fiscal expansion by core Europe, although those benefits are shown to be present only in a liquidity-trap environment.

lower euro interest rates, the resulting downward pressure on the international demand for the euro will likely bring about its nominal depreciation. The Expanded Asset Purchase Programme will likely help improve Greece's external competitiveness for two main reasons: One, offsetting deflationary pressures throughout the eurozone would suppress Greece's real exchange rate vis-à-vis its EU trading partners, especially while it remains under its stabilization program; and two, the program would reduce the borrowing costs of the Greek sovereign when it returned to the international markets. If nominal depreciation of the euro continues, it would help Greece in connection with its non-EU trading partners, but its competitiveness problem would remain, especially in connection with its EU trading partners. Solving the competitiveness problem would require a targeted approach to structural reforms.

There is a long history of linking debt relief with reforms that improve economic efficiency. As Barry Eichengreen, Peter Allen, and Gary Evans (2015) discuss, in 1991 Western governments, through the Paris Club, offered Poland a 30 percent cut in the present value of its debt in return for agreeing with the IMF on the terms of a structural adjustment program. Poland subsequently received a further 20 percent cut which, importantly, was contingent on fulfillment of the structural conditions of its IMF program (Greenhouse 1991). The politics are vastly different, but Greece, too, has been offered conditional debt relief: First, in the Eurogroup Statement of November 27, 2012 (Eurogroup 2012), and then, most recently, in the Eurogroup Statement of August 14, 2015 (Eurogroup 2015).[18] However, in

18. The Eurogroup Statement on Greece, November 27, 2012, first grants Greece relief of its debt in a number of ways, and then states: "Euro area Member States will *consider further measures and assistance, including inter alia lower co-financing in structural funds and/or further interest rate reduction* of the Greek Loan Facility, if necessary, for achieving a further credible and sustainable reduction of Greek debt-to-GDP ratio, when Greece reaches an annual primary surplus, as envisaged in the current MoU [Memorandum of Understanding], conditional on full implementation of all conditions contained in the programme, in order to ensure that by the end of the IMF programme in 2016, Greece can reach a debt-to-GDP ratio in that year of 175 percent and in 2020 of 124 percent of GDP, and in 2022 a debt-to-GDP ratio substantially lower than 110 percent" (Eurogroup 2012, p. 2). The Eurogroup Statement on Greece, August 14, 2015, states: "The Eurogroup considers the continued programme involvement of the IMF as indispensable and welcomes the intention of the IMF management to recommend to the Fund's Executive Board to consider further financial support for Greece once the full specification of fiscal, structural and financial sector reforms has been completed and once the need for additional measures has been considered and an *agreement on possible debt relief to ensure debt sustainability* has been reached" (Eurogroup 2015, p. 2; emphasis added).

this latest instance, relief is conditional on its being deemed necessary for debt sustainability.[19]

Debt overhang affects the policy space of the Greek government and, in addition, fuels adverse expectations through its effect on individuals' perceived wealth. Given the loss of wealth associated with the unprecedented contraction since 2010, such expectations have a strong impact on the economy. Because a sovereign can always walk away from a deal, especially if it is running surpluses, and refuse to service the outstanding debt, it makes intuitive sense for the creditors to provide incentives associated with the implementation of structural reforms in the form of debt relief.

Andreas Müller, Kjetil Storesletten, and Fabrizio Zilibotti (2015), motivated by the events surrounding the Greek and eurozone crisis, develop a theory of sovereign debt to examine the properties of the optimal dynamic contract between a planner and a sovereign when the country cannot commit to honoring its debt. Their theory's main implication for Greece is that at high debt levels the incentive to reform is reduced, because most of the benefit from reform will go to the creditors. The optimal program requires that whenever a credible default threat is on the table, the lenders should give in and improve the terms of the agreement for the debtor country by granting it higher consumption and a lower reform effort. In other words, the austerity program should be relaxed over time, whenever this is necessary, to avert the breakdown of the program. These results clearly bear upon the negotiations between Greece and its creditors.

V. Concluding Remarks

During most of the time since Greece's accession into the euro area, the Greek government collected less in taxes than it spent, as indicated by its increasing fiscal deficits as a share of GDP, which rose from 4.5 to 15.6 percent between 2001 and 2009. In addition, the Greek economy consumed more than it produced and had to import far more than it exported, as indicated by current account deficits as a share of GDP, which rose from 7.2 to 14.6 percent between 2001 and 2008. As a result, Greece experienced an increase in its external public debt as a share of GDP from

19. Debt sustainability, which debt relief is meant to ensure, remains a bone of contention between the IMF and the eurozone due to different definitions and policy objectives. Such differences have, however, been played down by the eurozone in recent informal statements. See Spiegel (2015).

103.7 percent in 2001 to 129.7 percent in 2009, in spite of generous help from the EU's structural funds.

The Economic Adjustment Programme has been a major "demand" force in the severe contraction since 2009, but there is also a "supply" force. Greece must further improve its competitiveness vis-à-vis its eurozone partners, and debt relief in and of itself cannot address this need. It requires a targeted approach involving structural reforms, especially reforms that improve competitiveness in the market for goods and services. Such reforms are necessary to make Greece more productive, help it attract investment, and aid it in developing forward-looking export industries. This will inevitably require deep restructuring of the economy, a process that typically follows crises and is, to some extent, already under way.

Reforms have effects over and above the impact of price and wage changes on unit labor costs. In Greece they are critical for another reason too, namely an adverse demographic outlook. Current population decline would make it harder for Greece to pay off its debt (Ioannides 2015). Reforms involve short-term costs and are thus painful, but they are necessary. Targeted smart reforms are under a nation's control, and it would be a tragedy if Greece did not undertake them, especially while under assistance. Debt relief alone will not solve the competitiveness problem, but if it is designed as an incentive to improve competitiveness, it could help.

ACKNOWLEDGMENTS We are truly grateful to the editors for their very helpful suggestions and to the copyeditors for manuscript preparation and meticulous fact and source checking, which improved the paper; to Anna Hardman, Gikas Hardouvelis, John Lipsky, Vasilis Nicoletopoulos, Enrico Spolaore, and Nikos Vettas for their generous comments on drafts prior to the conference; and to George Dimitriou for his help with data. We are entirely responsible for the content. Pissarides and Dimitriou's work was funded by the European Research Council through the project Employment in Europe at the University of Cyprus.

References

Alesina, Alberto, Silvia Ardagna, and Vicenzo Galasso. 2010. "The Euro and Structural Reforms." In *Europe and the Euro,* edited by Alberto Alesina and Francesco Giavazzi. University of Chicago Press.

Bi, Ran, Haonan Qu, and James Roaf. 2013. "Assessing the Impact and Phasing of Multi-Year Fiscal Adjustment: A General Framework." Working Paper no. 13/182. Washington: International Monetary Fund.

Blanchard, Olivier, Christopher J. Erceg, and Jesper Lindé. 2015. "Jump-Starting the Euro Area Recovery: Would a Rise in Core Fiscal Spending Help the Periphery?" Working Paper no. 304. Sveriges Riksbank.

Chen, Ruo, Gian Maria Milesi-Ferretti, and Thierry Tressel. 2013. "External Imbalances in the Eurozone." *Economic Policy* 28, no. 73: 101–142.

Cheptea, Christina, and Delia Velculescu. 2014. "A Disaggregated Approach to Prioritizing Structural Reforms for Growth and Employment." In *Jobs and Growth: Supporting the European Recovery.* Washington: International Monetary Fund.

Draghi, Mario. 2015. "Structural Reforms, Inflation and Monetary Policy." In *ECB Forum on Central Banking: Inflation and Unemployment in Europe, Conference Proceedings.* European Central Bank.

Dustmann, Christian, Bernd Fitzenberger, Uta Schönberg, and Alexandra Spitz-Oener. 2014. "From Sick Man of Europe to Economic Superstar: Germany's Resurgent Economy." *Journal of Economic Perspectives* 28, no. 1: 167–88.

Eble, Staphanie, Wojciech Maliszewski, Maral Shamloo, and Iva Petrova. 2013. "Greece: Selected Issues." Country Report no. 13/155. Washington: International Monetary Fund.

Eichengreen, Barry, Peter T. Allen, and Gary Evans. 2015. "Escaping the Greek Debt Trap." *Bloomberg View,* July 27.

Eurogroup. 2012. "Eurogroup Statement on Greece." November 27. http://www.consilium.europa.eu/workarea/downloadAsset.aspx?id=19043

———. 2015. "Eurogroup Statement on the ESM Programme for Greece." August 14. http://www.consilium.europa.eu/press-releases-pdf/2015/8/40802201771_en_635751847200000000.pdf

European Commission. 2010. "The Economic Adjustment Programme for Greece." Occasional Paper no. 61. Brussels.

———. 2013. "Shadow Economy and Undeclared Work." http://ec.europa.eu/europe2020/pdf/themes/07_shadow_economy.pdf

———. 2014. "The Second Economic Adjustment Programme for Greece: Fourth Review." Occasional Paper no. 192. Brussels.

———. 2015. "Memorandum of Understanding between the European Commission Acting on Behalf of the European Stability Mechanism and the Hellenic Republic and the Bank of Greece." Brussels.

Galenianos, Manolis. 2015. "The Greek Crisis: Origins and Implications." Research Paper no. 16. Athens: Crisis Observatory, Hellenic Foundation for European and Foreign Policy (ELIAMEP).

Greenhouse, Steven. 1991. "Poland Is Granted Large Cut in Debt." *New York Times,* March 16.

Grossman, Gene M., and Esteban Rossi-Hansberg. 2008. "Trading Tasks: A Simple Theory of Offshoring." *American Economic Review* 98, no. 5: 1978–97.

Hardouvelis, Gikas. 2015. "How to Undo the Damage in Greece." *New York Times,* July 6.

Hausmann, Ricardo. 2012. "Ireland Can Show Greece a Way out of the Crisis." *Financial Times,* February 8.

Hausmann, Ricardo, César A. Hidalgo, Sabastián Bustos, Michele Coscia, Alexander Simoes, and Muhammed A. Yildirim. 2013. *The Atlas of Economic Complexity: Mapping Paths to Prosperity.* MIT Press.

IMF (International Monetary Fund). 2010. "Will It Hurt? Macroeconomic Effects of Fiscal Consolidation." In *World Economic Outlook: Recovery, Risk, and Rebalancing,* October. Washington.

———. 2013. "Greece: Ex Post Evaluation of Exceptional Access under the 2010 Stand-By Arrangement." Country Report no. 13/156. Washington.

———. 2014. "Greece: Fifth Review under the Extended Arrangement under the Extended Fund Facility, and Request for Waiver of Nonobservance of Performance Criterion and Rephasing of Access; Staff Report; Press Release; and Statement by the Executive Director for Greece." Country Report no. 14/151. Washington.

———. 2015a. "Blanchard: Looking Forward, Looking Back." Interview with Olivier Blanchard. *IMF Survey Magazine,* August 31.

———. 2015b. "Where Are We Headed? Perspectives on Potential Output." In *World Economic Outlook: Uneven Growth—Short- and Long-Term Factors,* April. Washington.

Ioannides, Yannis M. 2015. "Why Productivity-Enhancing Reforms Will Help Greece Exit the Crisis and Usher in Long-Run Growth." Working Paper. https://sites.tufts.edu/yioannides/files/2015/01/IoannidesJan242015MegaronPaperFigs.pdf

Kent, Christopher, and John Simon. 2007. "Productivity Growth: The Effect of Market Regulations." Research Discussion Paper no. 2007-04. Reserve Bank of Australia.

Koske, Isabell, Isabelle Wanner, Rosamaria Bitetti, and Omar Barbiero. 2015. "The 2013 Update of the OECD's Database on Product Market Regulation: Policy Insights for OECD and Non-OECD Countries." Economics Department Working Paper no. 1200. Paris: OECD Publishing.

Lyberaki, Antigone, Costas Meghir, and Daphne Nikolitsas. Forthcoming. "Labour Market Regulation and Reform in Greece." In *Reforming the Greek Economy,* edited by Costas Meghir, Christopher Pissarides, Dimitri Vayanos, and Nikos Vettas. MIT Press.

Moreno-Badia, Marialuz, Bob Traa, and Delia Velculescu. 2009. "Greece: Selected Issues." Country Report no. 09/245. Washington: International Monetary Fund.

Müller, Andreas, Kjetil Storesletten, and Fabrizio Zilibotti. 2015. "Sovereign Debt and Structural Reforms." Discussion Paper no. 10588. London: Centre for Economic Policy Research.

National Bank of Greece. 2014a. "Greece Macro View: March 2014." Athens. https://www.nbg.gr/greek/the-group/press-office/e-spot/reports/Documents/Greece%20Macro%20View%20_March%202014%20Final.pdf

———. 2014b. "Greece Macro View: July 2014." Athens. https://www.nbg.gr/greek/the-group/press-office/e-spot/reports/Documents/NBG_Greece%20Macro%20Focus_Labour%20Market%20July14%20fin.pdf

———. 2015. "Greece Macro View: October 2015." Athens. https://www.nbg.gr/greek/the-group/press-office/e-spot/reports/Documents/Greece_MacroView_October2015fin.pdf

Nicoletti, Giuseppe, and Stefano Scarpetta. 2003. "Regulation, Productivity and Growth: OECD Evidence." *Economic Policy* 18, no. 36: 9–72.

OECD (Organization for Economic Cooperation and Development). 2013. *OECD Economic Surveys: Greece 2013.* Paris: OECD Publishing.

Papageorgiou, Dimitris, and Evangelia Vourvachaki. 2015. "The Macroeconomic Impact of Structural Reforms in Product and Labour Markets: Trade-Offs and Complementarities." Working Paper no. 197. Bank of Greece.

Pissarides, Christopher A. 2013. "Is Europe Working?" LSE Regius Professorship Inaugural Lecture, December 12. London School of Economics.

———. 2015. "Structural Perspectives on European Employment: The Role of Innovation and Growth." In *ECB Forum on Central Banking: Inflation and Unemployment in Europe, Conference Proceedings.* European Central Bank.

Rahman, Jesmin, and Tianli Zhao. 2014. "The Role of Vertical Supply Links in Boosting Growth." In *Jobs and Growth: Supporting the European Recovery.* Washington: International Monetary Fund.

Rehn, Olli. 2013. "Turning Germany's Surplus into a Win-Win for the Eurozone." Blog, November 11. European Commission. http://ec.europa.eu/archives/commission_2010-2014/blogs/rehn/turning-germanys-surplus-into-a-win-win-for-the-eurozone.html

Schwab, Klaus. 2012. "The Global Competitiveness Report 2012–2013." Geneva: World Economic Forum.

———. 2014. "The Global Competitiveness Report 2014–2015." Geneva: World Economic Forum.

Spiegel, Peter. 2015. "Klaus Regling Interview: The Annotated Transcript." Brussels Blog. *Financial Times,* October 2.

Vargas, Janos, Werner Roeger, and Jan in 't Veld. 2013. "Growth Effects of Structural Reforms in Southern Europe: The Case of Greece, Italy, Spain and Portugal." Economic Paper no. 511. Brussels: European Commission.

World Bank. 2013. "Doing Business 2014: Understanding Regulations for Small and Medium-Size Enterprises." Washington.

———. 2014. "Doing Business 2015: Going beyond Efficiency." Washington.

Zettelmeyer, Jeromin, Christoph Trebesch, and Mitu Gulati. 2013. "The Greek Debt Restructuring: An Autopsy." *Economic Policy* 28, no. 75: 513–63.

GENERAL DISCUSSION All four papers were the subject of the discussion that follows.

Gregory Mankiw opened with a question about one of the findings in Christopher House and Linda Tesar's paper. They had found that imposing a labor or consumption tax and putting the proceeds toward paying off foreign creditors would severely depress GDP and, therefore, not raise as much revenue as one might hope. But Mankiw thought the correct parameter for evaluating this would be the uncompensated elasticity of labor supply, which was either close to zero or even backward-bending in the long run, and as a result it should not have a large effect on GDP after all. He wondered if their finding differed from this because, rather than following the standard assumption that income effects are similar to or somewhat larger than substitution effects, as he assumed, they had treated substitution effects as larger than income effects. Or was it because of Keynesian effects from sticky prices and, if so, how long do those effects last?

In Mankiw's view, the reason one cannot simply tax labor or consumption in Greece to pay off all the creditors is that there are limits to how much one can apply such methods and not because of Keynesian or neoclassical effects found in the modeling, such as income and substitution effects. He thought it probably has more to do with the political instability and tax evasion that would result at a certain point.

Ben Friedman spoke up to comment on the political-economy implications of the Greek problem, including the threat it posed to the structure of the euro area. He was surprised none of the authors had mentioned the banks and the way they were bailed out. In his opinion, it was a great tragedy that the Europeans paid, and are still paying, a great price for the way they handled their bank bailout. In the United States, by contrast, the government let the banks absorb losses during the crisis and then recapitalized those that needed recapitalizing, including some very large ones such as Citibank and Bank of America. The Europeans shied away from that approach and instead moved many of the questionable debts, as soon as it became clear the Greeks might default, from the banks' balance sheets to those of the central banks. Friedman thought Carmen Reinhart and Christoph Trebesch were right in concluding that debt relief is what has been needed all along. Private sector lenders know how to handle the situation of borrowers being in trouble and figuring out what to do with those debts, even though it may be a messy solution—it is after all what bankers get paid to do. But when debts are on the balance sheets of the official lenders, one is stuck in the fiction that they must never accept a default.

If the upshot were just a matter of some governments having to take losses that they did not want to admit, that might not have been too serious, Friedman argued, but now it has reached the point at which it has affected the high politics of the European Union (EU). He offered by way of analogy the situation in the Americas today in which Argentina is in default to many lenders, many of them domiciled in the United States. One cannot easily imagine President Obama being asked what his opinion is on Argentinian debt—he would probably respond by asking reporters why they thought he should have an opinion on the subject at all—and it certainly does not affect the high politics between Argentina and the United States. By contrast, up until the refugee crisis hit Europe, Angela Merkel, the Chancellor of Germany, was unable to hold any press conference without being asked what she thought about the Greek crisis.

David Romer had three big-picture questions stemming from three of the papers. He commended Christopher House and Linda Tesar for their paper's narrow focus on the feasibility of tax and spending options for solving the debt problem, which he read as concluding that those options are inadequate due to what amount to leakages in dynamic scoring. His first question was why House and Tesar did not take the next step and examine alternative ways to solve the problem. Two methods occurred to him: debt write-downs, which Reinhart and Trebesch mentioned in their paper, and structural reform, which Yannis Ioannides and Christopher Pissarides underlined as an approach with potentially enormous value. If Greece could raise its growth through other means, it could solve the debt problem by making the denominator in its debt ratio bigger.

Second, he was curious how far Reinhart would be willing to take her policy prescriptions. Would she advocate abolishing foreign borrowing if, for example, she were the newly installed president of a Latin American country? Certainly higher Greek domestic saving would be great, but achieving that is extremely hard. One could imagine instead returning to the practices of the 1950s and 1960s, when there was little mobility in international capital. Is that what she would ideally like to see?

His third question was aimed at Beatrice Weder di Mauro and coauthor Julian Schumacher, who advocated that the International Monetary Fund (IMF) abolish its systemic exemption. While it seemed true that the IMF had set up an incredibly discretionary process and that there were big disadvantages of that, it also appeared that if the IMF had not been willing to break its rules and had allowed Greece to go into a disorderly default in the crisis period of 2010, the systemic consequences would have been enormous. Romer said he certainly would not have advocated that the IMF

simply abandon Greece to default. He wondered whether there might be a middle-ground approach that would allow for discretionary actions in exceptional circumstances without sacrificing all the benefits of a rule-based approach. His impression was that the authors had not fully articulated their position, and he wanted to hear more.

Bradford DeLong was struck by a finding in Reinhart and Trebesch's paper showing that, historically, real ex post returns on defaulted bonds were in the range of one to five percent, despite the losses due to haircuts and arrears. Notwithstanding Jeremy Bulow and Kenneth Rogoff's demonstrations that one should never do so, when crises come creditors somehow have enough control to squeeze the lemon hard, regardless of the excess burden in taxes and other costs imposed on the Greeks.

Echoing Romer's question to Reinhart, DeLong asked: Is it really the case that a country should never borrow in a currency it cannot print unless it happens to be Canada or Australia? And if so, should a country never let its firms borrow in a currency that it cannot print, because the private debt will be turned into a public debt during the crisis when everyone is looking to kick the can down the road? DeLong wondered whether alternative baselines were needed to assess this. Likewise, he wondered what would have been the macroeconomic consequences for Texas in the early 1990s had the U.S. government insisted that Texas reimburse the Resolution Trust Corporation for payments made to depositors in the Texas savings and loan crisis.

Kevin O'Rourke agreed with Romer's point that rules should be waived when there is a real systemic risk, but he also thought it should have a corollary requirement. In the case of deciding not to restructure a debt because of the systemic risk involved, should not the attendant cost be shared among all the members of the system that is being protected? He also agreed with DeLong's concern about the dangers that flow from private sector borrowing abroad. O'Rourke reminded everyone that in Europe, democracy resides at the level of the nation-state rather than at the trans-European level, so when a nation's sovereign decision-making is disrupted by debt crises, even temporarily, it creates a serious political problem. This led him to wonder whether, lacking a proper banking union, cross-border banking and lending should be reconsidered altogether.

He also pointed out a seeming contradiction between two of the papers. As he understood it, Weder di Mauro argued that the present value is what matters most when assessing debt levels, whereas Reinhart's paper alluded to the face value of a debt actually mattering most for economic

performance. Both views seemed plausible to him, so he would need to see more empirical evidence to choose between them.

Maurice Obstfeld found it striking that the Greek crisis started out as a debt crisis and only later evolved into a banking crisis centered on the relationship between the European Central Bank (ECB) and the Greek banks. He concurred with Weder di Mauro and Schumacher's observation that when the European Stability Mechanism (ESM) tries to evaluate systemic risk, it rates any risk of default as systemic, and added that this is mainly because a country in default might have to leave the eurozone due to the position of its banks. Obstfeld concluded that to make the ESM function credibly, for example so that collective action clauses can allow countries to actually default, one would need to have a complete banking union, something that is not present in the eurozone. The lack of such a union is an Achilles' heel in the eurozone arrangement today, but it seemed to him that a banking union could be established and would stabilize the eurozone, solving part of the ESM problem, even if it created political problems.

Ricardo Reis found it surprising that the central bank did not play a more central role in the presentations, especially as it concerns Greece. The Bank of Greece is the main source of outside funds for Greece, through its access to the ECB, its effect on interest rates, and in determining how much banks can raise. Moreover, as part of the euro system, the Bank of Greece is no longer able to choose monetary policy for Greece to accommodate fiscal policy changes there, so this key determinant of fiscal multipliers works quite differently from the model in the House and Tesar paper. Finally, while Weder di Mauro and Schumacher had focused on the public debt, Reis thought their analysis neglected the very large liabilities the Bank of Greece holds to the rest of the euro system, both through the Emergency Lending Assistance (ELA) program and through the target program.

It worried Reis that Greece's central bank still has in place capital controls with limits on deposits. A problem arises once many Greek citizens expect the country to exit the euro. For Greeks today, taking currency out of a bank means getting foreign currency, whereas leaving money in a bank risks seeing it decline in value after Greece leaves the euro. As a result, the optimal strategy for individuals is to get as many euros into their pockets as they can, even though this leaves a money multiplier equal to one—essentially eliminating the banking system. Whether the government raises taxes, forgives debt, reassesses sustainability, or engages in structural reforms, the dominant strategy for individual Greeks is to have no banking system working at all, clearly an unsustainable situation. It is

no longer simply a debt problem. Reis saw this as ultimately stemming from the way the ECB works, enabling an exchangeability between currency and deposits that has led to a bad speculative equilibrium.

Caroline Hoxby thought the focus needed to be on structural reform, something almost every observer seemed to agree is badly needed. The labor market is very distorted, deregulation is needed, monopolistic and oligopolistic practices such as governmental mispricing of purchases need to be ended—in short it seemed clear that many structural reforms are in order. But the ordinary Greek citizen does not seem to recognize this, and Greek leadership has been weak in supporting it and promoting it. In that regard, she wondered whether being integrated into the EU and the eurozone was helpful for making structural reform, or harmful. On the one hand, it certainly increases the pressure for structural reform because capital flows occur in competition with other European countries that have better institutions. But membership also decreases political pressure for structural reform because it makes it very easy for people to leave Greece and effectively live and work elsewhere.

Hoxby's comments prompted Donald Kohn to raise the issue of competitiveness. To remain in a currency union permanently and avoid serious economic pain, a country has to find a way to be competitive. What struck him about Ioannides and Pissarides's paper was the finding of a lack of price responsiveness as compared to wage responsiveness. Kohn wanted to know if the labor market was more competitive than the goods and services market, or if perhaps the declines in wages were concentrated in the public sector, where the prices cannot adjust. Are prices set in the EU common market framework? Whereas labor market costs do seem to be adjusting, prices in the product market do not, and this begs the question of what the right structural reforms might be.

Martin Baily added to the discussion of structural reform by noting a point raised at previous Brookings Panel conferences, namely that if structural reforms are in the first instance job destroying, in a Keynesian situation they can actually make things worse. Structural reforms that give new businesses a chance to open are certainly good, Baily said, but they will not raise employment when, for example, they allow big-box stores to arrive and drive out small businesses. Episodes like this might raise productivity but would not raise GDP. He felt structural reforms needed to solve the employment problem first before focusing on enhancing productivity. Baily also proposed that the problem of tax evasion in Greece is a priority. He recalled recent news reports citing high rates of uncollected statutory taxes, so that a better long-term solution than raising

tax rates would be to increase rates of compliance. The assessment of the two papers is that Greece's situation remains gloomy, and he found the solutions offered in the papers to be limited. Across Europe there is an aggregate demand problem, especially severe within Greece, and solutions should aim firstly at that problem. A contractionary fiscal policy—even in the form of collecting more uncollected taxes—may not be the right road in the short run.

Martin Feldstein agreed with Baily about the need for increased economic growth in Greece to finally resolve the problem, adding that nothing in the most recently negotiated deal suggested the potential to accomplish this. He has heard the same from European acquaintances he has spoken to about this—people who are much closer to the Greek crisis and its coverage.

Richard Cooper shifted the discussion to what he labeled the psychological side of structural reform. He noted Ioannides's report that not only ordinary Greeks but even Greek leaders felt unpersuaded about the need for reform, and had two questions for him. First, might the word *adjustment* be more useful in public debates, since the elements assumed to be in structural reform are understood very differently by different people? And second, what did Ioannides think about presenting the public with this simple proposition: One cannot consume more than one earns or produces without borrowing from some third party? Such a simple statement should be easy to get across to ordinary households. The Greek public appeared not to understand the gravity of running a current account deficit of 10 percent of GDP as well as a budget deficit in some years. Was this due to politicians' failure to communicate the issue to them? Or was it a failure of journalists and economists?

Ioannides replied by reminding everyone that the U.S. public did not seem to understand the same problem either. To that comment, Cooper answered, somewhat tongue in cheek, that the American public understood the problem of deficits quite well but simply dealt with them by thinking the U.S. government could borrow endlessly. Greece, by contrast, could not borrow, once the severity of its crisis was revealed, and that is the crux of the problem, Cooper said. Deficit adjustment then became necessary by simple arithmetic. To him, the real question was, why was that lesson not brought home to the Greek public? Or was the situation perhaps a collective gamble that Greece could dragoon the rest of the Europeans into lending more to them?

The authors of the four papers responded next, beginning with House. He noted that his and Tesar's paper had relied on certain parametric choices

in the modeling, including their use of Greek tax return data to measure the taxable base of labor income, which is not the same as the overall amount of labor income earned in the model. He added that Mankiw was correct in thinking that sticky prices and sticky wages are influencing the results, since price and wage rigidity are substantial in their model and dominate the model behavior, particularly in the short run. Concerning Greece's liability to outside creditors, it is measured in nominal terms, so production and consumption are not identical between Greece and the rest of Europe, and movements in relative terms of trade also interfere with the results.

In response to Romer's inquiry about ways to improve the Greek situation beyond tax and spending cuts, House said that in addition to debt write-downs, another tool is to allow Greece to delay repayment and to do so at below-market rates. This would mean setting up extremely favorable loan terms while keeping the face value of the debt fixed. It would create some breathing room although, admittedly, it would also run into a credibility problem.

Ioannides spoke next. He responded, first, to Hoxby's concern for the importance of structural reforms and the uncertainty whether achieving buy-in for them is easier or harder due to the free movement of labor that EU membership enables. He believed the net effect has been to make it easier, and certainly easier than under autarky, because membership has made the public more apt to learn from the successes of other countries, and in fact Greece's attachment to having a European identity has been a driver of much of the politics. Related to the last point, he mentioned the advantages of EU membership in enabling the importation of technology, research, and university education, with all their links to industry. In Greece today the linkages between industry and the universities are weak, so they represent an area that can be strengthened and should become a priority.

Price rigidity is indeed a problem, he added. Reforming the labor market was easier by comparison, and it should be recognized that workers today are receiving a fraction of what they were paid before the reforms. Product market reforms were in the agreements, as well, but implementing them has not been given priority and it is harder to carry out. Labor and product markets were certainly grossly noncompetitive up to the onset of the crisis in 2010.

Concerning Baily's point about the significance of uncollected taxes in Greece, Ioannides agreed that they represent a big problem. Businesses are also in arrears to the tax authorities, so the shockingly high number

cited earlier of 85 percent of taxes uncollected might not be far off if the sum total of arrears is included. More importantly, Ioannides said, one needs to know how much can realistically be collected. The tax authorities have made progress in identifying what a realistic compliance rate is. It turns out that if the compliance rate in Greece had been, prior to the onset of the crisis in 2010, the same as the average rate across the eurozone, there would have been no debt problem in the first place. The money that could have been collected was actually there, but the Greek government was just not doing its job in collecting.

Finally, he said, while the outlook of the Greek people concerning the seriousness of the debt is puzzling and even the more educated know little of the country's tragic history in this area, he remains hopeful that progress is being made. He believes that working with other Greek economists to educate the public and speak with the press are deeply important, and noted his own efforts to contribute to that as a blogger. A part of the public, especially in the unfolding political parties, understands the need for reform.

Reinhart then added her responses to the mix of author comments. Noting that a few speakers had raised the issue of banks and wondered why they were a focus of interventions, she pointed out that contagion was a serious concern. Although it was true that the Greek crisis had started as a debt crisis, not a banking crisis, that was not the case in the other countries, where the official sector took over privately held Greek debts to thwart the risk of contagion. She agreed with those who thought this approach has caused delays in the recovery, and not just in Greece, noting that in other crises public and private restructurings occurred much quicker.

The elephant in the room, Reinhart said, was the problem of Greece's external dependence, something Romer and DeLong raised. She believed it could be broken down into three problems to solve. First was the management of the government debt, which the examples of Mexico and Chile have shown to be a problem that can be dealt with. Second was the management of private sector debt, which is much trickier to resolve. She has long been concerned about surges in capital flows due to the procyclical nature of capital markets, which caused especially serious problems in Ireland, Spain, and Iceland. And the third aspect of external debt, which is an endemic problem in Greece, is the public's attitude toward reform measures. How do you convince the public that they will not have their wealth confiscated? Restoring public confidence is very difficult. In 1932, Greeks endured a forcible debt conversion from foreign currency deposits

to drachmas, so there is a precedent and the public has grounds to be worried that it will recur.

Finally, concerning making economic growth a priority, an issue Baily and Feldstein had raised, Reinhart pointed out that she and Trebesch had stressed the importance of haircuts for this same reason, to help restore growth. When one looks at restructuring episodes in Latin America, for example, one sees that the Baker Plan, which was to extend maturities, did not conclude the debt crisis there, whereas the Brady Plan, which included haircuts, did.

Weder di Mauro spoke last. Responding to O'Rourke's question whether it is the face value or the present value of debt that matters most, she said what matters for sustainability is the present value, since that is what defines a country's debt repayment burden. Unfortunately, what seems to matter in the headlines is the face value since most commentators are not aware of the highly concessional terms Greece enjoys. In the short run, Greek debt burdens are low. In the medium run, Weder di Mauro said, the repayments will increase and may breach the thresholds applied by the IMF and the ESM. This then raises questions about the need, the timing and the type of possible further restructuring of Greek debt held by the official sector. Restructuring by extending grace periods and maturities for European loans may decrease the medium-run repayment burden. However, it would extend the external dependence of Greece even further and increase the risk of repeated renegotiations and political clashes such as the one witnessed in 2015.

Concerning the role of the banks, Weder di Mauro pointed out that foreign banks took very high haircuts in the debt restructuring of 2012. There was a measure of coercion to achieve this: European governments leaned heavily on their banks and Greece retrofitted collective action clauses in debt contracts. Therefore, Greek debt held by European banks was not simply transferred to the ECB. However, over the course of 2015, the run on Greek banks' deposits had forced the ECB to extend ever more emergency liquidity assistance in order to prevent a shutdown of the Greek banking system and a de facto exit from the currency union.

Finally, Weder di Mauro considered Romer's doubts about her paper's proposal that the IMF end its policy of systemic exemption. Romer had suggested that if in 2010 the IMF had told Greece it could not grant access to a loan restructuring, Greece would have gone into a tailspin, but she believed that what would have happened is that the Europeans would have assisted in restructuring earlier than they did. She did not think Greece would have defaulted and exited the eurozone at that time.

In her view, at the very least the IMF and the ESM need to be clear what they mean by "systemic" risk and should also have to consider the alternative costs, that is, the costs of delayed restructuring and gambling for resurrection. It struck Weder di Mauro that a renewed debate is needed on the access conditions for international financial assistance both at the IMF and the ESM in the cases of doubtful debt sustainability.

JOSHUA K. HAUSMAN
University of Michigan

JOHANNES F. WIELAND
University of California, San Diego

Overcoming the Lost Decades? Abenomics after Three Years

ABSTRACT We review the recent performance of the Japanese economy under Abenomics, the set of economic policies begun by Prime Minister Shinzō Abe in 2012. We find that in 2014, Abenomics, and in particular expansionary monetary policy, continued to weaken the yen and raise stock prices. It also continued to generate positive inflation, though neither actual nor expected inflation is yet 2 percent. The real effects of Abenomics have been modest. Performance would have been better if not for two puzzles: The response of net exports to the weak yen was small, and there is little evidence that expansionary monetary policy had large effects on consumption.

Shinzō Abe took office as prime minister of Japan in December 2012 and embarked on a set of economic policies widely dubbed "Abenomics." Abe's economic program consisted of three arrows: (i) expansionary monetary policy, (ii) expansionary fiscal policy, and (iii) structural reforms. Under Governor Haruhiko Kuroda's leadership, the Bank of Japan has vigorously pursued expansionary policy. But fiscal policy, while initially expansionary, turned contractionary in April 2014 when the consumption tax was raised from 5 to 8 percent. And while progress has been made on some structural reforms, such as electricity deregulation, corporate governance, and female labor force participation, many of Abe's reform promises remain unimplemented (IMF 2015b). We therefore focus on monetary policy, the arrow of Abenomics that is both the most novel and the most fully implemented.

In April 2013, the Bank of Japan embarked on a program of "quantitative and qualitative easing," aiming to reach 2 percent inflation in two years (Bank of Japan 2013a). To achieve this goal, between 2012Q4 and 2015Q1 it

increased the monetary base from 25 percent of GDP to 57 percent of GDP. In the process, it accumulated 128 trillion yen of Japanese government bonds, equal to more than 25 percent of GDP.

In section I, we review the effects of Abenomics, and these monetary actions in particular, on intermediate indicators. Building on the analysis in our previous paper (Hausman and Wieland 2014),[1] we show that expansionary monetary policy continued to weaken the yen and raise stock prices in 2014. Yet effects on nonfinancial variables were muted. Inflation expectations from market participants and professional forecasters remain roughly one-half to one percentage point below the Bank of Japan's 2 percent target. Actual headline and core inflation are also still well below 2 percent. We argue that this persistent low expected inflation largely reflects the imperfect credibility of the 2 percent inflation target, although we cannot rule out some role for adaptive expectation formation and backward-looking price-setting behavior.

In section II, we consider the response of output to Abenomics. Between 2012Q4 and 2015Q2, annualized GDP growth was 0.9 percent; when measured per person ages 15 to 64 (the working-age population), it was 2.4 percent.[2] A comparison with the rest of the world during the same period suggests that this was a success. For example, between the end of 2012 and 2015Q2, annualized GDP growth per working-age person was 1.8 percent in the United States and 1.1 percent in Germany. Nevertheless, relative to professional forecasts, Japan's performance has been disappointing. Output in 2015 is likely to be at least a percent lower than that forecast in October 2012, before Abenomics began.[3]

Performance would have been better if not for two puzzles: weak consumption and weak net exports. Despite a 1-percentage-point decline in the real interest rate, consumption has been flat during the Abenomics years. To better understand this, we use the Japanese Family Income and Expenditure Survey to investigate how expansionary monetary policy is affecting different types of households. The results are puzzling, with monetary policy showing no visible effects on consumption among those households expected to benefit most, namely, net debtors and the young. In

1. For other recent evaluations of Abenomics and quantitative easing in Japan, see Patrick (2014) and Ito (2014).

2. All data are as of August 28, 2015. See the online appendix for information on sources. Online appendixes for papers in this volume may be found on the *Brookings Papers* web page (www.brookings.edu/bpea) under "Past Editions."

3. This assumes that actual 2015 output growth is equal to 1 percent, the Consensus Economics forecast made in April 2015.

contrast, the April 2014 consumption tax increase had large effects on the consumption of all types of households. The story of flat consumption in Japan may therefore be one in which expansionary monetary policy had relatively little positive effect while contractionary fiscal policy had large negative effects.

A further mysterious factor behind slow output growth is a large increase in real imports. Since Abenomics began, real imports have risen by more than 10 percent, despite flat consumption and a weakening yen. We discuss three popular hypotheses—a decline in the relative price of imports, an increase in energy import demand, and an increase in foreign electronics demand—but find all of them to be either unsupported by the data or too small to explain the size of the import increase. In our view, the increase in imports remains a puzzle.

In section III, we turn to the outlook for future output and consumption in Japan. Consensus forecasts are for the level of GDP over the next five years to be nearly the same as that forecast in October 2012, before Abenomics began. This is largely because the path of Japanese consumption is now forecast to be below that expected in October 2012. That in turn is consistent with a larger-than-expected negative effect of the consumption tax and the lack of progress in making structural reforms.

We concluded in our previous paper on the subject (Hausman and Wieland 2014) that the first arrow of Abenomics, expansionary monetary policy, most likely passed a cost-benefit test. This remains our conclusion. The magnitude of the benefits is uncertain, but for the reasons detailed in that paper, the costs are likely small. We end this paper with suggestions for how the Bank of Japan might provide additional stimulus to the economy.

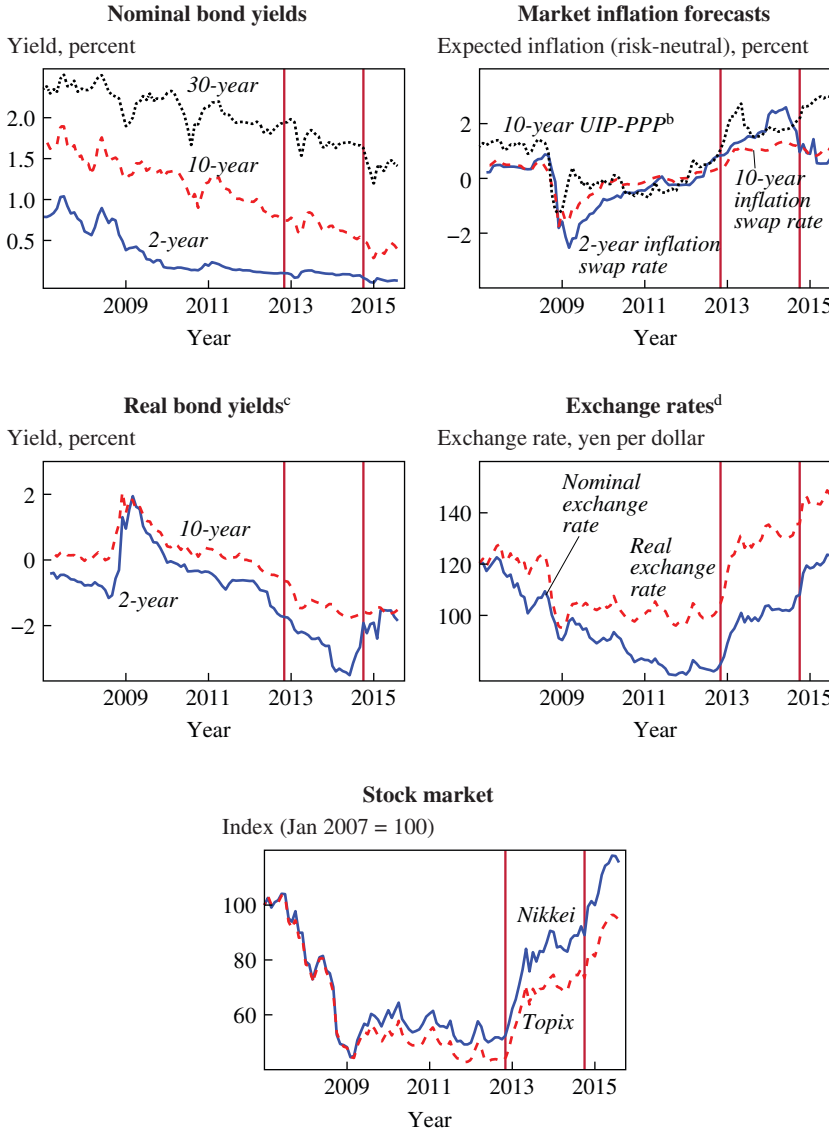
I. Intermediate Indicators

The ultimate goal of Abenomics is to raise output, but its effects on financial markets and inflation are also of interest. The response of these intermediate indicators sheds light on the mechanisms through which expansionary monetary policy is—and is not—affecting the economy.

I.A. Financial Markets

Abenomics has continued to have large effects on financial markets. Figure 1 shows updated versions of the financial market figures in our previous paper on this subject (figures 2 through 6 there; see Hausman and Wieland [2014]). Financial market developments have generally continued along their early 2014 paths. The two vertical lines in each panel correspond to November

Figure 1. Abenomics' Financial Market Effects, 2007–15[a]



Sources: Bloomberg, Bank for International Settlements, and Yahoo! Finance. See online data appendix for more details.

a. These figures are an update of figures 2–6 in Hausman and Wieland (2014). In all panels, the Abenomics period begins in November 2012, indicated by the first vertical line. The second vertical line denotes October 2014, when quantitative easing was expanded.

b. The UIP-PPP measure of inflation expectations is calculated using the uncovered real interest rate parity condition and U.S. TIPS. For details, see Krugman (2013) and Hausman and Wieland (2014).

c. Real bond yields are calculated as the difference between nominal bond yields and inflation swap rates.

d. The nominal exchange rate is yen per dollar. The real exchange rate is the broad BIS index, indexed to equal the nominal exchange rate in January 2007.

2012, when then candidate Shinzō Abe made clear his economic policy intentions, and to October 2014, when the Bank of Japan expanded its quantitative and qualitative easing program, raising the targeted annual increase in the monetary base from 60–70 trillion yen (12–14 percent of 2014 GDP) to 80 trillion yen (16 percent of GDP) (Bank of Japan 2014).

During 2014, the most dramatic financial developments occurred in the value of the yen and in Japanese stock prices. The yen weakened from 79 per dollar in October 2012 to 102 per dollar in March 2014 and then to 123 per dollar in August 2015. This nominal exchange rate movement was largely reflected in Japan's trade-weighted real exchange rate. According to the broad Bank for International Settlements (BIS) index, the real trade-weighted yen weakened 44 percent between October 2012 and July 2015. In July 2015, the real trade-weighted yen was weaker than at any time since 1982.[4] Stock prices also continued to rise rapidly. From October 2012 to March 2014, the broad Topix index rose 62 percent; between March 2014 and August 2015 it rose a further 36 percent.

Of course, the coincidence between these asset price movements and expansionary monetary policy alone is no proof that the movements were caused by monetary policy. The best evidence that the policy was a cause comes from movements in asset prices on the day of significant monetary policy announcements. We documented in the earlier paper (Hausman and Wieland 2014) that declines in nominal interest rates, declines in the value of yen, and increases in the stock market all coincided with news of expansionary policy. These effects are consistent with time-series evidence on the effects of quantitative easing in Japan (Ito 2014). Further evidence comes from the financial market reaction to the announcement of the expansion of quantitative and qualitative easing on October 31, 2014. On that day, 30-year bond yields fell 5 basis points, the yen weakened 2.8 percent against the dollar, and the Topix stock market index rose 4.3 percent.[5]

I.B. Inflation

A primary goal of Abenomics, particularly for its monetary arrow, is to end Japan's 15 years of deflation. So far, it has succeeded. The upper-left

4. This statement is based on the BIS *narrow* trade-weighted index, since the *broad* trade-weighted index begins only in 1994.

5. The interpretation of these movements is complicated by the fact that on the same day (October 31, 2014), Japan's Government Pension Investment Fund announced that it would be purchasing more Japanese and foreign stocks instead of Japanese bonds (Kitanaka, Nozawa, and Nohara 2014). The decline in bond yields on this day, however, suggests that the monetary policy announcement had larger financial market effects than the pension fund decision.

panel of figure 2 shows three measures of prices in Japan. In each measure, the effect of the 3-percentage-point increase in the consumption tax in April 2014 is obvious. But even apart from that tax increase, prices have generally risen.

However, the Bank of Japan has not achieved its stated goal of 2 percent inflation. From July 2014 to July 2015, the overall CPI rose 0.2 percent, while the CPI excluding food and energy rose 0.6 percent. We saw in the upper-right panel of figure 1 that market inflation expectations generally remain below 2 percent, and the four other measures shown in the upper-right panel of figure 2 confirm that there was little increase in inflation expectations during 2014. Firm inflation expectations, as measured by the Bank of Japan's Tankan survey,[6] and 1- and 10-year inflation expectations from Consensus Economics forecasts all remain below 2 percent.[7] Interestingly, there is no evidence that Japanese households expected deflation before or after Abenomics began; according to the Bank of Japan Opinion Survey, in the two years before Abenomics began (September 2010–September 2012), household inflation expectations averaged 3.7 percent. This fits with international evidence suggesting that households and small businesses are ill informed about inflation and monetary policy (Kumar and others 2015; Binder 2014).

THREE POSSIBLE MECHANISMS OF INCOMPLETE ADJUSTMENT Any or all of three possible mechanisms are likely driving the incomplete adjustment of expected inflation toward the 2 percent target: (i) backward-looking price setting, (ii) adaptive expectations (slow updating), and (iii) imperfect credibility. To better understand which of these factors is quantitatively most important, we conduct the following exercise: First, we estimate a Phillips curve for Japan following Olivier Coibion and Yuriy Gorodnichenko (2015b) in order to gauge the amount of backward-looking price setting. We use inflation forecasts and output gap data[8] to estimate a new Keynesian Phillips curve with a fraction β_1 of backward-looking firms,

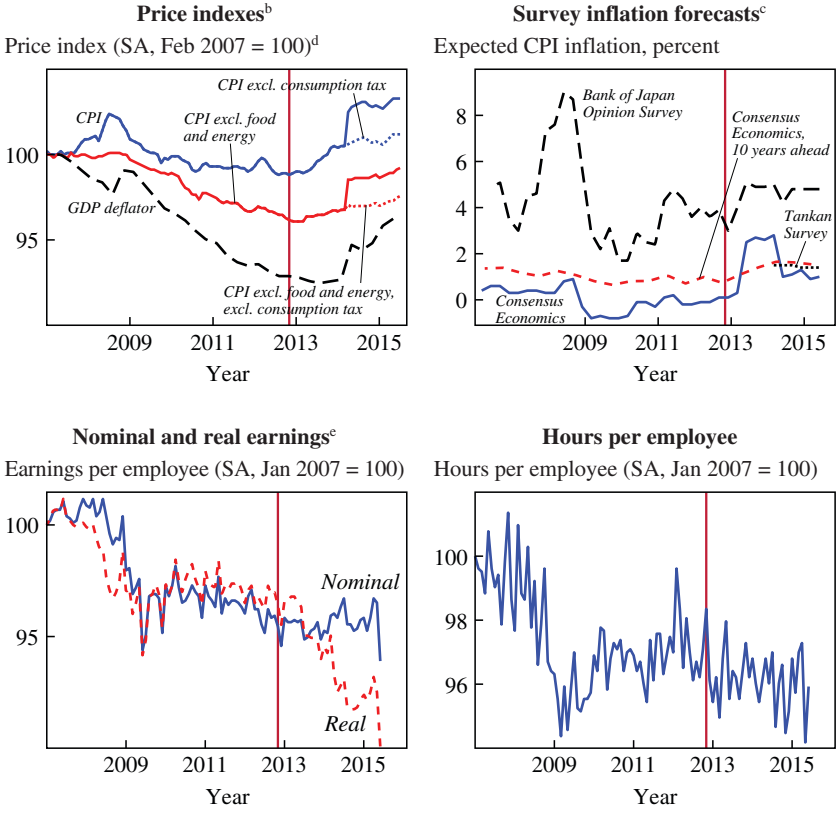
$$\pi_t - E_t\pi_{t+1} = \beta_1(\pi_{t-1} - E_t\pi_{t+1}) + \beta_2 x_t + \varepsilon_t. \quad (1)$$

6. The Tankan survey began to ask about inflation expectations only in March 2014.

7. The hump in 1-year inflation expectations in 2013 and 2014 reflects the influence of the April 2014 consumption tax increase.

8. We use the International Monetary Fund's data on the output gap from its April 2015 *World Economic Outlook* (IMF 2015a). This is not inconsistent with our argument in Hausman and Wieland (2014) that this measure of the output gap underestimates the possible effect of monetary policy on output over the long run; for the Phillips curve, what is relevant is potential output in the short run.

Figure 2. Japanese Actual and Expected Inflation, Wages, and Hours per Employee, 2007–15[a]



Sources: Japanese Statistics Bureau, Consensus Economics, Bank of Japan, and Japanese Ministry of Health, Labour and Welfare. See online data appendix for more details.

a. In all panels, the Abenomics period begins in November 2012, indicated by the vertical line.

b. Direct effects of the consumption tax are excluded from the CPI by assuming that the consumption tax raised 12-month headline inflation by 1.9 percentage points and 12-month headline inflation excluding food and energy by 1.5 percentage points in April 2014, and by 2.1 percentage points and 1.7 percentage points, respectively, from May 2014 through March 2015. These are the figures suggested by the Bank of Japan. See http://www.stat.go.jp/english/data/cpi/report/2014np/pdf/fu8.pdf.

c. Inflation forecasts are 1-year-ahead CPI forecasts, unless otherwise noted.

d. The GDP deflator is measured quarterly, and is indexed to 2007Q1 = 100.

e. Real earnings are nominal earnings deflated by the CPI, excluding imputed rent, but including the consumption tax.

Table 1. Phillips Curve Estimates[a]

Independent variables	*(1) OLS*	*(2) IV*[b]
$\pi_{t-1} - E_t\pi_{t+1}$	0.22	0.11
	(0.15)	(0.18)
Output gap x_t	0.14***	0.13
	(0.053)	(0.091)
p value[c]	0.38	0.59
F statistic		9.32
R^2	0.18	0.12
No. of observations	27	24

a. Each column is a regression estimation of equation 1; the dependent variable is inflation minus expected inflation. Column 1 is the ordinary least squares approach, and column 2 is the instrumental variables approach. Newey-West standard errors in parentheses. Statistical significance indicated at the ***1 percent, **5 percent, and *10 percent levels.

b. Instruments are a lag of the output gap x_{t-1}, and lagged forecasts $\pi_{t-1} - E_{t-1}\pi_{t+1}$.

c. Testing for a vertical Phillips curve, which is the *t* test on β_3 in the equation $\pi_t - E_t\pi_{t+1} = \beta_1(\pi_{t-1} - E_t\pi_{t+1}) + \beta_2 x_t + \beta_3 E_t\pi_{t+1} + \varepsilon_t$.

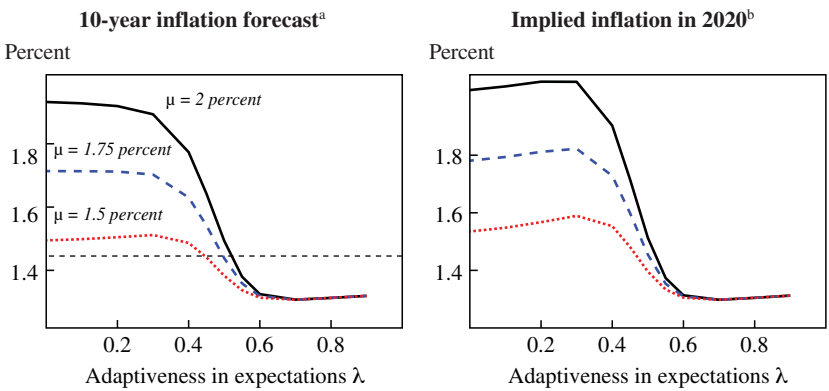
Note that backward-looking price setting β_1 is distinct from adaptive expectations, the latter of which are contained in $E_t\pi_{t+1}$. By measuring inflation expectations directly, we capture the adaptiveness of forecasts. This allows us to take the expectations formation process as given and to then isolate the amount of backward-looking price setting necessary to explain the observed persistence of inflation.

We estimate this equation by ordinary least squares (OLS) and instrumental variables on annual data from 1989 to 2015, where the instruments are a lag of the output gap x_{t-1} and lagged forecasts $\pi_{t-1} - E_{t-1}\pi_{t+1}$. We use the instrumental variable approach, standard in this literature (Coibion and Gorodnichenko 2015b), to avoid the bias that would result when contemporaneous supply shocks move the output gap and expected inflation in opposite directions. The structure imposes a vertical long-run Phillips curve, a restriction not rejected by the data. Table 1 shows results; they suggest that backward-looking price setting is small, with β_1 below 0.25.

Second, given equation 1, we solve for the expected path of inflation using output gap forecasts and a terminal condition that inflation reach a target μ (in this case 2 percent) in 2030.[9] This corresponds to the (credible)

9. We linearly extrapolate the *World Economic Outlook's* 2020 output gap forecast to reach zero in 2030. Results are not sensitive to this assumption, since the projected output gap in 2020 is small. Results are also not sensitive to extending the time horizon at which the inflation target becomes credible.

Figure 3. Phillips Curve Simulation Results



a. Plots implied 10-year inflation forecasts from solving the estimated Phillips curve from equation 1 and expectations formations from equation 2 for combinations of credible long-run inflation μ and adaptiveness λ. The horizontal dotted line is the 10-years-ahead inflation forecast of 1.45 percent from Consensus Economics.
b. Plots implied inflation rate in 2020 for combinations of μ and λ. See note a.

expected rate of inflation in the very long run. We allow for expected inflation in the Phillips curve to be partially adaptive:

$$E_t \pi_{t+1} = \lambda \pi_{t-1} + (1 - \lambda) \pi_{t+1}, \tag{2}$$

where λ indexes adaptiveness, and π_{t+1} is the solution to equation 1. Rational expectations correspond to the case $\lambda = 0$, and fully adaptive expectations correspond to $\lambda = 1$.[10] By iterating on equations 1 and 2 until convergence, we can determine what combinations of long-run actual inflation μ and what degree of adaptiveness λ can rationalize both the market and professional long-run inflation forecasts. We use the OLS estimates to parameterize the Phillips curve (column 1 of table 1), but these results are very similar to the instrumental variable estimates.

In the left panel of figure 3, we show the implied 10-year inflation forecasts for combinations of long-run inflation μ and adaptiveness λ. The inflation forecasts follow an inverse-S shape in λ. This is because the

10. Sticky information corresponds to a generalization in which $\lambda_t = (1 - \theta)^t$, and θ is the fraction of agents updating information every year. The literature suggests that θ is roughly between 0.68 and 0.94 (Mankiw and Reis 2002; Mankiw, Reis, and Wolfers 2004; Coibion and Gorodnichenko 2015a), which implies $\lambda \approx 0$ today given that the 2 percent target was announced in 2013. Thus, we view this case as being roughly captured by the $\lambda = 0$ calibration.

importance of λ for the forecast increases exponentially until it completely dominates the forecast. Intuitively, more adaptive expectations directly keep inflation low by increasing the weight on low past inflation. But since rational price setters correctly forecast this influence, they will also expect lower inflation, which further reduces price pressure today. Formally, inflation is a weighted average of initial inflation in 2015 and terminal inflation in 2030 (abstracting from the output gap), $\pi_t = g_t\pi_{2015} + (1 - g_t)\mu$, where g_t is given by a recursion.[11] The recursion implies that the weight on past inflation increases rapidly for intermediate values of λ. For $\lambda = 0$, the weight on past inflation in 2020 is $g_{2020} = 0.0005$, but it rises to $g_{2020} = 0.279$ for $\lambda = 0.3$, and increases steeply to $g_{2020} = 0.976$ for $\lambda = 0.5$. Consequently, we observe a sharp drop in inflation forecasts in the range $\lambda \in [0.3, 0.5]$. As this parameter increases further, the weight on past inflation becomes so large that future inflation is almost irrelevant, and the paths converge for different levels of long-run inflation μ.

Our simulation suggests that rationalizing the long-run 1.45 percent inflation forecast from Consensus Economics requires either a large degree of adaptiveness in expectations ($\lambda \approx 0.45$ to 0.55) or that forecasters believe long-run inflation μ will be only 1.5 percent, or some combination of those two possibilities. These high values for λ imply that, 5 years from now, the weight on the initial inflation target g_{2020} will range from 0.921 to 0.994. The higher inflation target is therefore almost irrelevant for price setting, even in 2020. Consequently, our simulations imply that inflation in 2020 will still be less than 1.6 percent if the entire deviation of current expected inflation from the 2 percent target is explained by adaptive expectations (see the right-hand panel of figure 3). This exercise suggests to us that a lack of credibility, that is, a belief that long-run inflation will fail to reach 2 percent, likely plays an important role, since full credibility implies such an extreme degree of sluggishness in inflation adjustment.

NOMINAL WAGE GROWTH Along with inflation expectations, we argued previously (Hausman and Wieland 2014) that nominal wage growth would be a critical indicator of Abenomics' success. This is because nominal wage growth is both a cause and an effect of inflation expectations, and also because real wages are likely to be an important determinant of consumption. Here the data continue to be disappointing. The lower-left panel of figure 2

11. Specifically, the weight on initial inflation is $g_t = \Pi_{i=2016}^{t} f_i$, where f_i is determined by the recursion $f_{2029} = \beta_1$, and $f_{i-1} = \dfrac{\beta_1 + (1 - \beta_1)\lambda}{1 - (1 - \beta_1)(1 - \lambda) f_i}$ for $i = 2017, \ldots, 2028$.

shows nominal earnings per person in the Japanese economy since 2007, and reveals no obvious increase in those earnings after Abenomics begins. Consequently, the recent increase in prices (upper-left panel of figure 2) has meant a steady decline in real earnings. From 2014Q2 to 2015Q2, real CPI-deflated[12] earnings per employee fell 1.4 percent; earnings per hour fell 0.9 percent. Cumulatively, over the three years from 2012Q2 to 2015Q2, real earnings per employee fell 5.0 percent; per hour, they fell 3.7 percent.[13]

There are likely three principal reasons why Abenomics has yet to translate into higher nominal wages, let alone higher real wages. First, the decline in real wages reflects in part a compositional effect due to a rising share of lower-paid part-time employment (Aoyagi and Ganelli 2015; Sommer 2009). But even among both full-time and part-time workers, real wages fell during Abenomics. Between 2012Q2 and 2015Q2, the real hourly earnings of full-time workers fell 3.5 percent, and those of part-time workers fell 0.8 percent. A shift in the composition of employment toward part-time work does not alone explain the decline in real wages.

A second reason Abenomics has failed to bring wage growth is the small change in inflation expectations, in particular expectations among firms regarding prices for their own products. In addition to asking firms about their CPI forecasts, the Tankan survey asks firms what they expect to happen to prices for their own output. In the June 2015 survey, firms expected to raise their own output price by an average of 0.9 percent over the next year. It is perhaps unsurprising, then, that firms are reluctant to pay higher nominal wages. An exception to this reluctance is large exporters, which have benefited from the weak yen, making it easier for them to grant wage increases. At Toyota, for instance, workers received a 3.2 percent increase in monthly pay during the spring 2015 Shuntō (annual spring wage negotiations).[14] However, the aggregate data show that this example is not representative.

12. Following the convention of the Monthly Labour Survey from the Japanese Ministry of Health, Labour, and Welfare, we report real wages as nominal earnings deflated by the CPI excluding imputed rent.

13. These data are from the Monthly Labour Survey, Japan's establishment employment survey. The figures from this survey include only "regular" employees; these are employees working more than one month or who were employed for the majority of the previous two months, including part-time employees. The sample covers private, nonagricultural industries. For more details, see http://www.mhlw.go.jp/english/database/db-slms/dl/slms-01.pdf and http://dsbb.imf.org/pages/sdds/DQAFBase.aspx?ctycode=JPN&catcode=WOE00.

14. On Toyota's profits, see Kubota (2015b). The wage figure excludes bonuses (Nakamichi and Kubota 2015).

A third reason for Abenomics' failure to lift wages is that the labor market may still be weak. For those age 15 to 64, the employment-to-population ratio steadily rose to nearly 73 percent in 2014[15] and, in absolute terms, Japanese unemployment is low: 3.3 percent in July 2015. Relative to the average unemployment rate in the 1980s of 2.5 percent, however, current unemployment in Japan is high. Moreover, the rise in participation and decline in unemployment have not been accompanied by an increase in monthly hours (see the lower-right panel of figure 2): Between 2012Q2 and 2015Q2, average monthly hours worked per full-time employee were unchanged, while average hours for all employees fell 1.4 percent.

The disappointing response of wages to Abenomics has led to political pressure and tax incentives for firms to increase wages. Both Prime Minister Abe and Governor Kuroda have pressured firms to raise wages.[16] In addition to this moral suasion, in 2013 the Abe administration introduced a tax credit for firms indexed to their wage bill.[17] Whatever the economic merit of such policies, however, they have not yet led to real wage growth.

II. Output

Macroeconomic theory suggests that the monetary arrow will contribute to higher output by lowering real interest rates and weakening the yen, thus raising consumption, investment, and net exports.[18] We first discuss overall growth before turning to the behavior of consumption and net exports in more detail.

15. Most of this increase came from a rise in the female employment-to-population ratio from 61 percent in 2012 to 64 percent in 2014. For more on this trend, see Posen (2014).

16. See Nakamichi and Fujikawa (2015) and Aoyagi and Ganelli (2015). A historical analogy to the efforts of the Abe administration to persuade firms to raise wages may be found in the efforts of U.S. presidents Herbert Hoover and Franklin Roosevelt to persuade firms to raise wages during the Great Depression (Rose 2010; Cole and Ohanian 2004). These policies remain controversial, with the benefits of higher inflation and inflation expectations (Eggertsson 2012) needing to be weighed against the costs of labor market distortions (Cole and Ohanian 2004; Friedman and Schwartz 1963; Cohen-Setton, Hausman, and Wieland 2015).

17. See Aoyagi and Ganelli (2015) and the October 4, 2013, issue of KPMG's *Japan Tax Newsletter* (https://www.kpmg.com/Jp/en/knowledge/article/japan-tax-newsletter/Documents/stimulate-business-investment-20131004E.pdf).

18. For monetary policy to have real effects, there have to be slack resources in the economy. In Hausman and Wieland (2014), we argue that this is the case and that official estimates of the output gap underestimate the scope for demand-based policies.

Table 2. Macro Summary Statistics, Japan and the United States, 1974–2014[a]

Period	*Real GDP growth (percent change)*	*Unemployment rate*	*CPI inflation*	*Money market interest rate*
1974–92 average	4.0	2.3	4.8	6.8
1993–2007 average	1.1	4.1	0.1	0.6
2008–12 average	–0.2	4.6	–0.2	0.2
2013	2.3	4.0	1.4	0.1
2014	–0.8	3.6	0.4[b]	0.1

	Real GDP growth per person (ages 15–64)		*Multifactor productivity (percent change)*		*Employment-to-population ratio (ages 15–64)*	
Period	*Japan*	*U.S.*	*Japan*	*U.S.*	*Japan*	*U.S.*
1974–92 average	3.1	1.5	—	—	67.4	68.1
1993–2007 average	1.4	1.9	0.7	1.1	69.3	72.5
2008–12 average	0.5	0.1	0.3	0.7	70.8	67.8
2013	3.7	2.1	1.5	0.4	71.7	67.4
2014	0.8	2.0	—	—	72.7	68.1

Sources: Japanese Cabinet Office, International Monetary Fund, Organization for Economic Cooperation and Development, Japanese Statistics Bureau, and Federal Reserve Bank of St. Louis, FRED database. See data appendix for more details.

a. This table is a reproduction of table 1 from Hausman and Wieland (2014) with updated data. All figures are percentages except where indicated otherwise. Growth rates are year-over-year except for 2013 and 2014, which are Q4-over-Q4 for GDP and inflation.

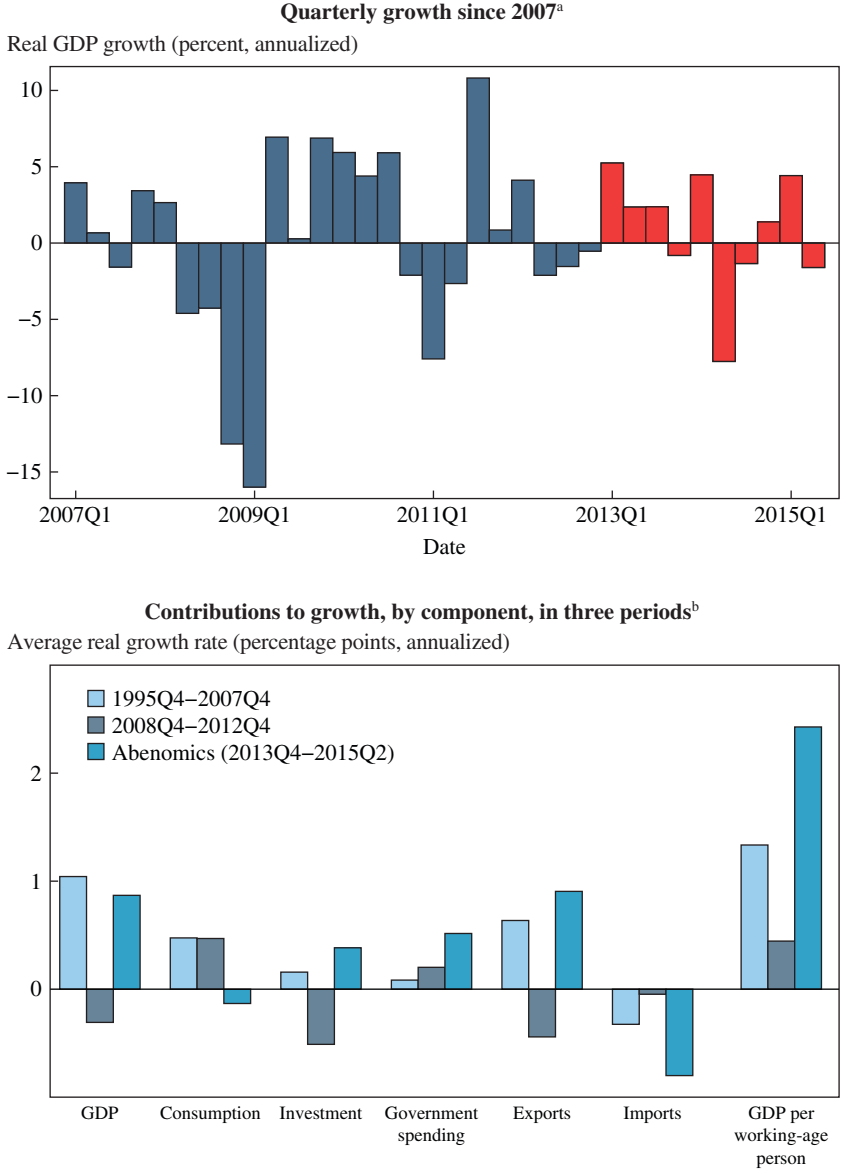
b. This excludes the direct effect of the April 2014 increase in the consumption tax from 5 to 8 percent. Including the consumption tax, CPI inflation was 2.5 percent. See http://www.stat.go.jp/english/data/cpi/report/2014np/pdf/fu8.pdf.

II.A. Overall Growth

Unfortunately, the overall growth effects of Abenomics so far appear to be small. Table 2 reproduces table 1 in our earlier paper (Hausman and Wieland 2014) and adds two lines showing the performance of the Japanese economy in 2013 and 2014. The table shows that relative to Japan's experience during its two lost decades of the 1990s and 2000s, performance in 2013 was excellent while that in 2014 was mediocre. Real GDP grew 2.3 percent between 2012Q4 and 2013Q4 (upper panel), which translated to 3.7 percent growth per working-age person (lower panel). This growth was more rapid than that in Japan during the boom decades of the 1970s and 1980s. Unfortunately, growth turned negative in 2014. Real GDP in Japan in 2015Q2 was 2.2 percent above its 2012Q4 level. Real gross domestic income was 2.8 percent above its 2012Q4 level.

The upper panel of figure 4 provides a more fine-grained perspective; it shows quarterly GDP growth at an annual rate in Japan since 2007. One

Figure 4. Japanese GDP Growth, 2007–15



Sources: Japanese Cabinet Office; Organization for Economic Cooperation and Development. See online data appendix for more details.

a. Shows annualized quarter-over-quarter real GDP growth since 2007. Quarters since Abenomics began are marked in red.

b. Provides a comparison of annualized contributions to real GDP by component during Abenomics (2012Q4–2015Q2), the lost decade excluding the Great Recession (1994Q4–2007Q4), and the Great Recession (2007Q4–2012Q4), as well as annualized working-age adjusted GDP growth. Contributions are calculated as in Japan's national accounts. See http://www.esri.cao.go.jp/en/sna/data/sokuhou/files/2011/qe114_2/pdf/kiyoe2.pdf

sees a clear reversal of progress after the consumption tax was raised in April 2014 from 5 to 8 percent. This panel also highlights the high volatility of measured Japanese GDP growth,[19] which makes it difficult to draw strong conclusions from one or even two or three quarters of growth. We therefore provide in the lower panel of figure 4 an alternative way of assessing recent performance that aggregates across several quarters. It shows contributions to growth under Abenomics compared with contributions to growth during Japan's lost decade (1995–2007) and the Great Recession (2008–12).

The differences between the left-most set and right-most set of bars in this lower panel (as well as in the results in table 2) indicate that any comparison of current Japanese economic performance to performance in the past or in other countries ought to make an adjustment for Japan's unusual demographics. Over the Abenomics period (2012Q4–2015Q2), the total Japanese population fell 0.5 percent and the working-age population (ages 15–64) fell 3.8 percent. Consequently, while overall GDP rose 2.2 percent between 2012Q4 and 2015Q2, GDP per capita rose 2.7 percent and GDP per working-age person rose 6.2 percent. By comparison, over this period, U.S. GDP per capita rose 3.9 percent and GDP per working-age person rose 4.5 percent. However, this comparison may exaggerate Japan's performance, for two reasons. First, some of growth early in Abenomics was likely bounce-back from a late 2012 recession (Hausman and Wieland 2014). Second, the working-age population adjustment ignores rising labor force participation among those of working age (see table 2) as well as a growing population between ages 65 and 74,[20] many of whom work (Kawata and Naganuma 2010). Still, the demographically adjusted figures show that Japan's performance under Abenomics has been far from dismal.

What is disappointing is the poor performance of consumption and net exports, as well as the shortfall of growth relative to what was forecast before and after Abenomics began. In 2015Q2, consumption was 0.6 percent below its level in 2012Q4. And during Abenomics, most of the positive contribution to growth from exports has been canceled out by a negative contribution from imports. We turn next to an analysis of this puzzling behavior of consumption and net exports.

19. Over the 20-year period from 1995 through 2015Q2, the standard deviation of quarterly (nonannualized) GDP growth in Japan was 1.1 percent, while that in the U.S. was 0.6 percent.

20. See National Institute of Population tables at http://www.ipss.go.jp/p-info/e/psj2012/PSJ2012-02.xls.

II.B. Evidence from the Cross-Section of Consumption Expenditures

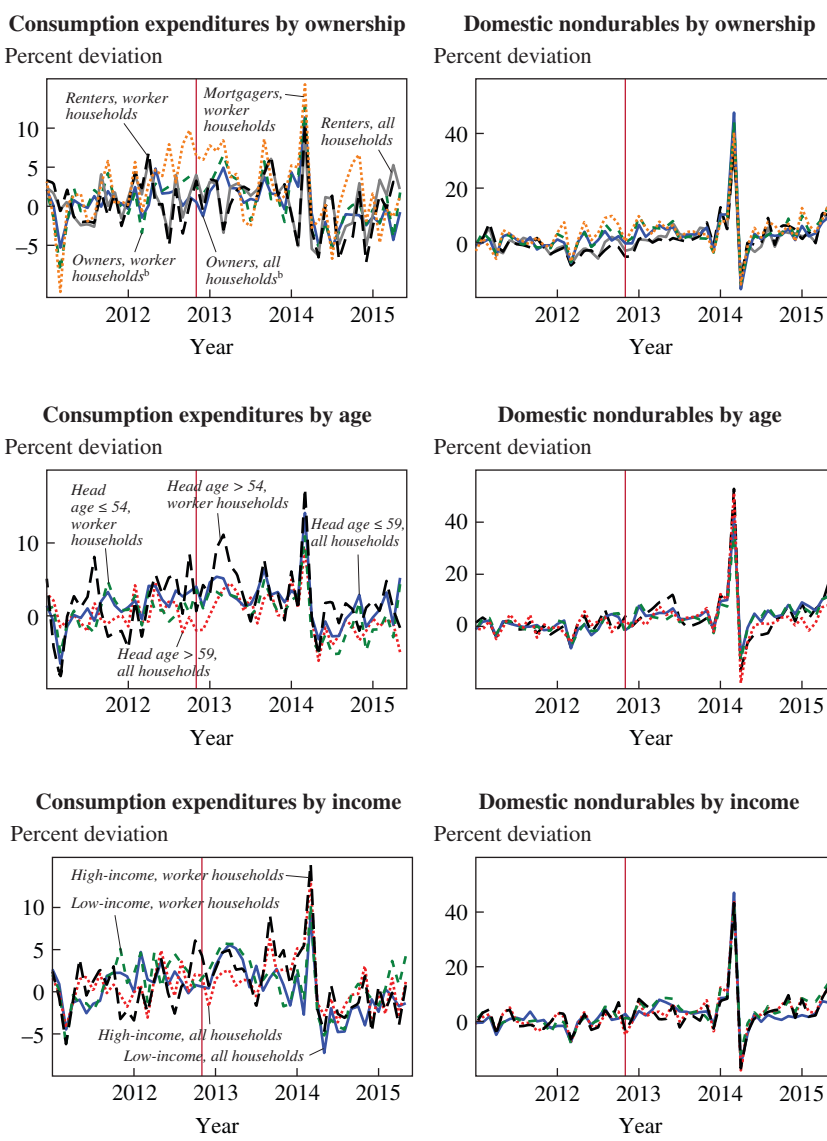
We now examine cross-sectional household expenditure data to learn more about the behavior of Japanese consumption under Abenomics. Like the aggregate time series, the cross-sectional data suggest that expansionary monetary policy has had little effect on consumption. We use the Japanese Family Income and Expenditure Survey, a survey of approximately 9,000 Japanese households. The Japanese Statistics Bureau publishes a breakdown of survey household consumption by home ownership status, age bins, and income quintiles.[21] We deflate these series by the CPI and seasonally adjust each series using an X-12 ARIMA(1,1) model with 12 monthly dummies.

The monetary policy arrow of Abenomics should have differential effects on these groups. First, higher expected (and actual) inflation constitutes a transfer from which mortgagors ought to benefit relative to renters and homeowners (Eggertsson and Krugman 2012; Cloyne, Ferreira, and Surico 2015). We also expect older households to be less willing to intertemporally substitute given finite horizons (Del Negro, Giannoni, and Patterson 2015), to be more likely to be creditors, and to be less likely to benefit from any labor market improvements under Abenomics. Thus, we would expect their responses to be muted relative to the responses of younger households. The breakdown by income is more ambiguous: higher-income households may have a greater ability to intertemporally substitute (McKay, Nakamura, and Steinsson 2015; Werning 2015), but poorer households may increase consumption more when income or credit supply grows.

Figure 5 plots real total consumption expenditures and domestic nondurable consumption expenditures along these dimensions relative to their 2011 log levels. We do not observe strong trends that would confirm the cross-sectional predictions raised in the previous paragraph. The trend of consumption by mortgagors looks quite similar to the trends among renters and owners; the consumption trend among the elderly is similar to that among the young; and the consumption trends of the rich and the poor likewise are similar. In part, this inference is a product of the noise in the consumption series. The repeated cross-sectional nature of the data does not allow us to filter any noise. A detailed study of the microdata might be better able to reveal differential effects of monetary policy, but with the data at hand we fail to see much evidence for large effects. This may

21. These data are only accessible from the Japanese version of the website at http://www.e-stat.go.jp/SG1/estat/eStatTopPortal.do. We are grateful to Hiroshi Matsushima for help with translation.

Figure 5. Household Consumption by Home Ownership, Age, and Income, 2011–15[a]



Source: Family Income and Expenditure Survey, Japanese Statistics Bureau.

a. Plots real total consumption expenditure and domestic nondurable consumption expenditures relative to their 2011 log levels. Before plotting, data are deflated by the CPI and seasonally adjusted using an X-12 ARIMA(1,1) model with 12 monthly dummies. The Abenomics period begins in November 2012, indicated by the vertical line.The full sample includes all households with at least two members; a "worker household" is one with at least two members whose head is employed as a wage earner in a public or private enterprise (excluding executives).

b. Owners include mortgagors.

be particularly surprising, since the cleaning up of Japanese firms' balance sheets over the last decade should have increased the traction of monetary policy (Kuttner and Posen 2001; Koo 2003; Wieland and Yang 2015).

In contrast to the absence of evidence for effects of monetary policy on consumption, the effects of the 3-percentage-point increase in the consumption tax in April 2014 are clear. In advance of the consumption tax, consumption boomed. It then plummeted. That the consumption tax had large intertemporal effects whereas monetary policy did not may be surprising, but this response is in fact consistent with standard models. In online appendix B, we consider a problem of consumption choice over storable and nonstorable nondurable consumption goods as in the work by Robert Barsky, Christopher House, and Miles Kimball (2007).[22] In this model, an anticipated consumption tax raises current consumption by lowering the real interest rate (the intertemporal price of consumption). Given the discrete nature of the consumption tax, the decrease in the real interest rate just before the tax hike is large relative to storage costs. This gives rise to a discrete increase in consumption expenditures. By contrast, if monetary policy causes only a smooth change in prices and the real interest rate, then it may not be optimal for consumers to discretely adjust their expenditures.

Likely adding to the effects of the April 2014 consumption tax increase was the fact that at that same time consumers were expecting the consumption tax to rise by a further 2 percentage points in October 2015.[23] (In fact, after the poor performance of the Japanese economy in the second and third quarters of 2014, the Abe administration postponed the October 2015 consumption tax increase to April 2017.) This added to the incentive to buy storable goods in advance of the April 2014 tax increase.

II.C. The Puzzling Behavior of Net Exports

The performance of net exports under Abenomics has also been disappointing. Between 2012Q4 and 2015Q2, real exports grew 15 percent and real imports grew 12 percent.[24] While one might have hoped for (even) stronger export growth, the larger mystery is why real import volumes have grown so rapidly despite a weaker yen and slow real output growth. Had

22. For another model of the effects of the consumption tax, see Cashin and Unayama (forthcoming).

23. We are grateful to Takashi Unayama for making this point to us.

24. There is a break in the Japanese balance of payments data due to item reclassifications at the start of 2014, with some effect on the real export and import data (Bank of Japan 2013b). In online appendix A.1, we provide further details and argue that adjusting for reclassification does not change the broad story of rising real export and import volumes.

import volumes remained flat, Japanese real GDP would have grown by as much as 3.8 percent rather than 2.2 percent since the end of 2012. This upper bound assumes a complete substitution of imports with domestic goods, but even for intermediate rates of substitution growth would have been noticeably faster.

We do not have a fully convincing explanation for the recent rise in import volumes. But we can rule out three hypotheses. First, one might wonder if Japan's import prices have in fact increased following the 56 percent depreciation of the yen against the dollar between October 2012 and August 2015. Perhaps the combination of falling commodity prices and pricing-to-market for other imports meant the yen depreciation was not associated with higher import prices. However, the data suggest otherwise. Measured by the import price deflator, between 2012Q4 and 2015Q2 import prices rose 7.0 percent. To be sure, this is far less than the yen depreciated, but it is nonetheless substantial.

Second, one might be tempted to ascribe the increase in import volumes to the substitution of fossil fuels for nuclear power in the aftermath of the 2011 Fukushima disaster. Fossil fuel imports did increase after Japan shut down its nuclear reactors, but this increase occurred before Abenomics began in late 2012. Between the first half of 2012 and the first half of 2015, the quantity of petroleum and liquid natural gas imports actually fell, while imports of coal rose by less than 4 percent. A further problem for this hypothesis is that it cannot explain why import volumes of services rose even more rapidly than those of goods during Abenomics; between 2012Q4 and 2015Q2, real imports of goods rose 9.7 percent while real services imports rose 22.9 percent.

Third, the International Monetary Fund (2014) suggests that the increase in real imports reflects growing Japanese demand for foreign electronics. Like the energy hypothesis above, this cannot explain the rise of service imports. But aside from this, the limited data available suggest it is an incomplete explanation. The yen value of Japanese imports of computers and phones (broadly defined[25]) rose by 1.2 trillion between the first half of 2012 and the first half of 2015. Had this rise not occurred, overall nominal Japanese imports would have risen 17.9 percent rather than 19.5 percent. Therefore, even with falling import prices for electronics and rising import prices for other goods and services, it is difficult to see how this story could account for very much of the increase in real Japanese import volumes.

25. We include computers and parts, semiconductors, audio and visual equipment, and telephony and telegraphy in this calculation.

III. Medium- to Long-Run Outlook

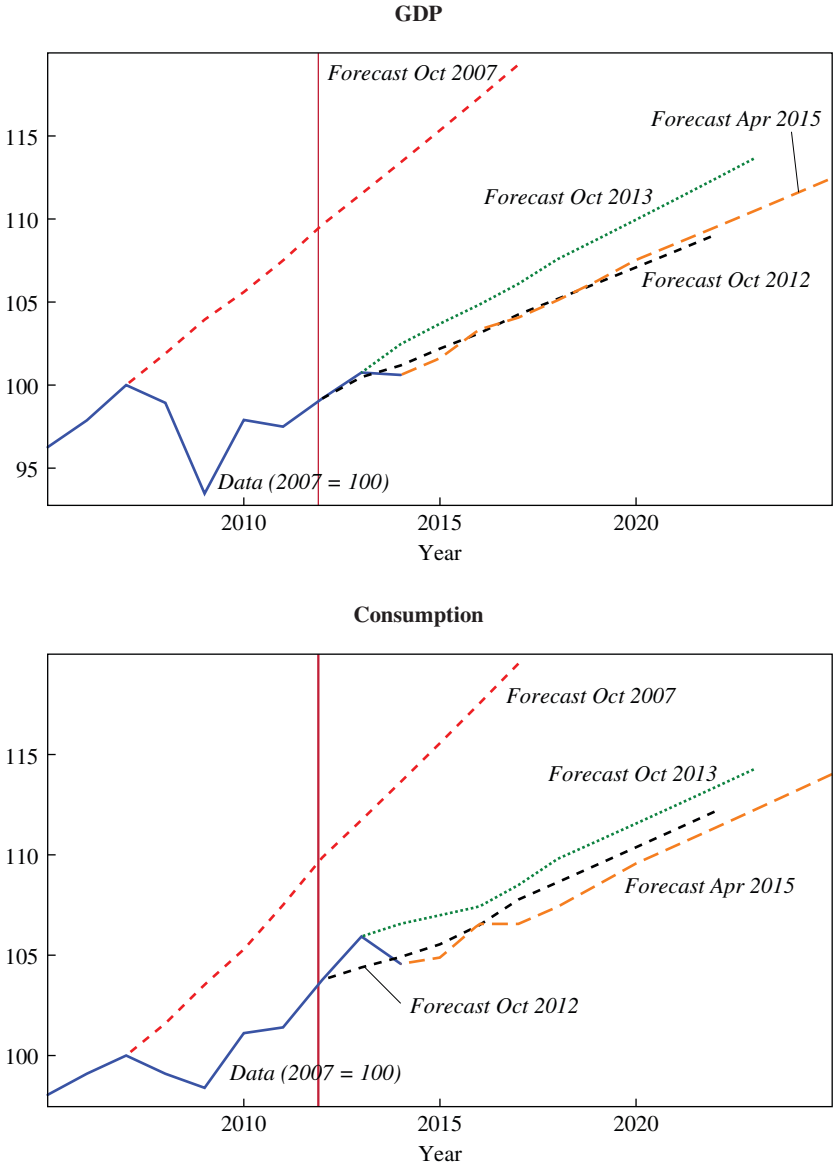
Japan's lackluster economic performance over the past 2 years is a reminder of the difficulty of macroeconomic forecasting. Both professional and model-based forecasts have been, to varying degrees, too optimistic. Thus we are now more pessimistic than we were 18 months ago about the long-run output effects of Abenomics.

Figure 6 updates figure 11 from our previous paper (Hausman and Wieland 2014), showing long-run professional forecasts from Consensus Economics for the level of real Japanese output and consumption. In our previous paper, we compared the forecast made in October 2013 to that made in October 2012, with the increase in the level suggesting real gains from Abenomics. Unfortunately, as the solid line shows, actual output and consumption have been below the level forecast in October 2013. Perhaps more troubling, long-run forecasts have reverted to their pre-Abenomics level in the case of output, and are below their pre-Abenomics level in the case of consumption. Importantly, there has been little change to Japanese demographic forecasts since 2006, so the change in output and consumption forecasts shown in figure 6 cannot be directly explained by demographic surprises.[26] This is worrisome both because the forecast may be correct and because it is an indicator of lackluster growth expectations.

When the Abenomics policies were first announced, there were at least two reasons to be more optimistic. First, given that the policies reduced real interest rates by roughly one percentage point, conventional new Keynesian models suggested output gains in the 5 to 10 percent range (Hausman and Wieland 2014). Second, a natural historical analogy for Abenomics is to the regime change engineered by Franklin Roosevelt in spring 1933 (Kuroda 2013; Romer 2014). In the four years after 1933, U.S. real GDP growth averaged 9.4 percent.

26. In 2006, the Japanese National Institute of Population and Social Security research forecast that the total Japanese population in 2015 would be 126.3 million and the working-age population (ages 15–64) would be 77.3 million. The latest projections (from 2012) are for these figures to be 126.6 million and 76.8 million (see note 20 for link to online tables). Of course, despite the accuracy of these demographic forecasts, it is possible that they were not fully incorporated into macro forecasts. The 2007 GDP forecast shown in the upper panel of figure 6 provides possible, but unclear, evidence for this. Using the 2006 population forecasts, it implies annualized per capita GDP growth from 2007 to 2015 of 1.9 percent and per working-age person growth of 2.8 percent. Using actual data on the size of the Japanese labor force, the 2007 GDP forecast implies annual growth of 2.0 percent per member of the labor force. These are optimistic forecasts but not obviously extreme. For example, U.S. GDP per capita grew on average at 2.0 percent per year between 1870 and 2014 (Jones 2015).

Figure 6. Actual and Forecast Output and Consumption, 2005–25[a]



Sources: Japanese Cabinet Office and Consensus Economics.

a. These figures are an update of figure 11 in Hausman and Wieland (2014). The Abenomics period begins in 2012, indicated by the vertical line.

There are three reasons why these model- and history-based predictions may thus far have been wrong. First, slow growth might be due primarily to the consumption tax increase. Unfortunately, quantifying the negative effects of the consumption tax on output is difficult since estimates of the tax multiplier in Japan vary widely (Kuttner and Posen 2001; Keen and others 2011). However, the observed large negative effects of the consumption tax in 2014 provide evidence supporting the high multipliers reported by Kenneth Kuttner and Adam Posen (2001). If the tax multiplier is large, fiscal consolidation will continue to depress Japanese output in the medium run.[27]

Second, Abenomics might be affecting the economy only with a long lag. Estimates for conventional monetary policy suggest that the peak effect on output is reached after 18 to 24 months (Christiano, Eichenbaum, and Evans 1999; Romer and Romer 2004). Regardless of whether one views Abenomics as having started with Abe's political campaign in November 2012 or with the announcement of qualitative and quantitative easing in April 2013, an 18- to 24-month lag suggests the peak effects ought to have already occurred. For Abenomics, however, the net export response may be unusually slow. For example, in July 2015—more than two years after the yen significantly weakened—Honda and Nissan announced that they would make a substantial shift toward producing cars in Japan for export.[28] This suggests that credibility of continued expansionary policy may be an important determinant of the net export response under unconventional monetary policy.

A third and final possibility is that the new Keynesian model and the 1933 analogy might be poor guides to the current Japanese macroeconomy. Recent events in Japan align with a growing literature suggesting that the new Keynesian model may exaggerate the output effects of forward guidance (Del Negro, Giannoni, and Patterson 2015; McKay, Nakamura, and Steinsson 2015). In our previous paper (Hausman and Wieland 2014), we documented that the change in the real interest rate in Japan since 2012 has been much smaller than that which occurred in the United States after 1933. Furthermore, lower real interest rates in the United States occurred along with other policy changes, such as financial reform,

27. In its April 2015 *World Economic Outlook,* the International Monetary Fund (2015a) predicts that Japan's structural budget deficit as a percent of potential GDP will decline by slightly more than one percentage point in both 2015 and 2016 and by roughly half a percentage point in 2017 and 2018.

28. See Kubota (2015a).

public works programs, and new regulations for businesses.[29] Moreover, the regime change in 1933 occurred after a precipitous fall in output and prices.

These three possible factors suggest that Abenomics, as is, is unlikely to substantially raise long-run output in Japan. However, the "as is" qualifier is important, since neither the monetary arrow nor the structural arrow appears to (yet) be fully credible policies.

IV. Credibility and Alternative Policies

We documented in sections I.A and I.B that most indicators of inflation expectations in Japan remain well below 2 percent, and we argued that this likely reflected imperfect credibility. One possible explanation for this lack of credibility, discussed in our previous paper (Hausman and Wieland 2014), is that observers doubt that there is political will to continue large-scale quantitative easing. Another possibility is that observers doubt the effectiveness of quantitative easing.

Insofar as there are doubts about the political will to achieve 2 percent inflation, it was unfortunate that the Bank of Japan's expansion of quantitative easing in October 2014 passed with only a 5-4 vote. By contrast, the decisive victory of Abe's Liberal Democratic Party in the December 2014 parliamentary elections may have increased confidence that monetary easing will continue. And in spring 2015, two members of the Bank of Japan's policy board (Ryuzo Miyao and Yoshihisa Morimoto) stepped down and were replaced by Yutaka Harada and Yukitoshi Funo in a transition that likely increased support for further easing.[30] The appointment of these new members did not lead to large changes in inflation expectations, however, suggesting that there are other sources of the credibility problem.

Given that quantitative easing has not (yet) produced actual or expected 2 percent inflation, the Bank of Japan could consider following the examples of Denmark, Switzerland, and the eurozone in paying negative nominal interest rates on reserves. Willem Buiter (2009) and Miles Kimball (2013) provide a discussion of the potential benefits of this policy. In the United

29. For more on policies and outcomes in the United States after 1933, see Romer (1992), Temin and Wigmore (1990), and Fishback (2008), among many others.

30. See Nakamichi and Ito (2015) and Ito and Nakamichi (2015). Harada is an economist who wrote a book entitled *Reflationalist Economics That Saved Japan* [in Japanese] (Harada 2014). Funo is a former Toyota executive; since his appointment, he has spoken publicly in favor of the 2 percent inflation target (Fujioka and Hidaka 2015). Harada replaced Ryuzo Miyao, who voted in favor of the October 2014 expansion of quantitative easing, while Funo replaced Yoshihisa Morimoto, who voted against this further easing (Bank of Japan 2014).

States, there is a concern that negative nominal interest rates could cause a run on systemically important money market funds by forcing them to "break the buck." In Japan, however, the importance of money market funds is negligible, less even than in Europe. As of 2014, money market shares amounted to $2.5 trillion in the United States, €427 billion ($467 billion) in the eurozone, and ¥14 trillion ($113 billion) in Japan. As a share of broad money this amounts to 18.3 percent in the United States, 4.1 percent in the eurozone, and 1.1 percent in Japan. This suggests that paying a negative interest rate on reserves might be a practical policy in Japan.

Negative nominal rates are only one of many alternative policies available to the Bank of Japan. For instance, as discussed by Lars Svensson (2003), the central bank could deliberately weaken the yen and peg the yen at a weak value. While net exports have not responded strongly to the recent yen depreciation, it is plausible that a peg could increase these effects by persuading firms of the weak yen's permanence. Such a peg might also improve the credibility of the 2 percent inflation target. A practical difficulty is that exchange rate policy falls within the scope of the Ministry of Finance rather than the Bank of Japan, so more explicit cooperation between them would be required.

We are hesitant to comment on more nonstandard proposals, such as money-financed government expenditures or money-financed fiscal transfers. Our analysis above suggests uncertainty about what macroeconomic model applies to Japan. This in turn implies uncertainty about how alternative policies would affect inflation and output.

V. Conclusion

In this paper we reviewed recent developments in Japan. Our analysis of Abenomics, and its monetary policy in particular, suggests that its real effects have so far been small despite intermediate indicators, such as the real interest rate and the real exchange rate, moving in an expansionary direction.

We focused less on the third arrow, structural reforms, in part because many reforms remain unimplemented and in part because professional forecasts suggest few further reforms will occur. Since late 2013, growth forecasts have declined (see figure 6) while inflation expectations have slightly risen (see the upper-right panel of figure 2). This is the opposite of the pattern one would expect if structural reforms were viewed as becoming more likely. In many standard macroeconomic models, structural reforms would raise growth expectations while lowering inflation expectations.

Thus, one could interpret stable inflation expectations and declining growth expectations as evidence of a declining probability that there will be further structural reforms.

That is the bad news. Good news may come in the form of the Trans-Pacific Partnership. This trade agreement would mean the liberalization of Japan's highly protected agricultural sector with a resulting large decline in food prices (Posen 2014). Furthermore, there is no lack of positive structural reforms available to Japan. For instance, the International Monetary Fund (2015b) estimates that reforms to increase the labor force participation of women and older persons could raise potential GDP growth by 0.25 percentage point per year. And Jamal Haidar and Takeo Hoshi (2015) provide many examples of high-return, low-cost reforms to regulations on new and existing businesses. While such reforms would undoubtedly be politically difficult, without making them Abenomics might have little impact on long-run growth.

ACKNOWLEDGMENTS We are grateful for insightful suggestions from the editors, David Romer and Justin Wolfers, and from our discussant, Adam Posen. We also thank Andrea De Michelis, Sohei Kaihatsu, Mitsuru Katagiri, Takushi Kurozumi, John Leahy, Kohei Shintani, Sayuri Shirai, Takashi Unayama, Kenichirou Watanabe, and seminar participants at the Bank of Japan and the Ministry of Finance for helpful comments. This work also benefited from the comments of our discussants, Michael Hutchison and Takatoshi Ito, on our prior paper (Hausman and Wieland 2014) at the 2014 Stanford Summer Juku. Hiroshi Matsushima provided superb research assistance.

References

Aoyagi, Chie, and Giovanni Ganelli. 2015. "Labor Market Reform: Vital to the Success of Abenomics." In *Can Abenomics Succeed? Overcoming the Legacy of Japan's Lost Decades,* edited by Dennis P. J. Botman, Stephan Danninger, and Jerald Alan Schiff. Washington: International Monetary Fund.

Bank of Japan. 2013a. "Introduction of the 'Quantitative and Qualitative Monetary Easing.'" Announcement by the Policy Board of the Bank of Japan. https://www.boj.or.jp/en/announcements/release_2013/k130404a.pdf

———. 2013b. "Revision of Balance of Payments–Related Statistics in Japan." BOJ Reports & Research Papers. https://www.boj.or.jp/en/research/brp/ron_2013/data/ron131008a.pdf

———. 2014. "Expansion of the Quantitative and Qualitative Monetary Easing." Announcement by the Policy Board of the Bank of Japan. http://www.boj.or.jp/en/announcements/release_2014/k141031a.pdf

Barsky, Robert B., Christopher L. House, and Miles S. Kimball. 2007. "Sticky-Price Models and Durable Goods." *American Economic Review* 97, no. 3: 984–98.

Binder, Carola. 2014. "Fed Speak on Main Street." Working Paper. https://sites.google.com/site/carolabindereconomics/research

Buiter, Willem H. 2009. "Negative Nominal Interest Rates: Three Ways to Overcome the Zero Lower Bound." *North American Journal of Economics and Finance* 20, no. 3: 213–38.

Cashin, David, and Takashi Unayama. Forthcoming. "Measuring Intertemporal Substitution in Consumption: Evidence from a VAT Increase in Japan." *Review of Economics and Statistics.*

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 1999. "Monetary Policy Shocks: What Have We Learned and to What End?" In *Handbook of Macroeconomics* 1A, edited by John B. Taylor and Michael Woodford. Amsterdam: North-Holland.

Cloyne, James, Clodomiro Ferreira, and Paolo Surico. 2015. "Monetary Policy When Households Have Debt: New Evidence on the Transmission Mechanism." Paper prepared for the Summer Institute 2015 Monetary Economics Workshop. Cambridge, Mass.: National Bureau of Economic Research.

Cohen-Setton, Jérémie, Joshua K. Hausman, and Johannes F. Wieland. 2015. "Supply-Side Policies in the Depression: Evidence from France." Working Paper no. W2015-08. Berkeley: University of California, Berkeley Economic History Laboratory.

Coibion, Olivier, and Yuriy Gorodnichenko. 2015a. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review* 105, no. 8: 2644–78.

———. 2015b. "Is the Phillips Curve Alive and Well After All? Inflation Expectations and the Missing Disinflation." *American Economic Journal: Macroeconomics* 7, no. 1: 197–232.

Cole, Harold L., and Lee E. Ohanian. 2004. "New Deal Policies and the Persistence of the Great Depression: A General Equilibrium Analysis." *Journal of Political Economy* 112, no. 4: 779–816.

Del Negro, Marco, Marc Giannoni, and Christina Patterson. 2015. "The Forward Guidance Puzzle." Staff Report no. 574. Federal Reserve Bank of New York.

Eggertsson, Gauti B. 2012. "Was the New Deal Contractionary?" *American Economic Review* 102, no. 1: 524–55.

Eggertsson, Gauti B., and Paul Krugman. 2012. "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach." *Quarterly Journal of Economics* 127, no. 3: 1469–513.

Fishback, Price V. 2008. "New Deal." In *The New Palgrave Dictionary of Economics,* 2nd edition, edited by Steven N. Durlauf and Lawrence E. Blume. London: Palgrave Macmillan.

Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1867–1960.* Princeton University Press.

Fujioka, Toru, and Masahiro Hidaka. 2015. "BOJ Must Achieve 2% Price Target, New Board Member Funo Says." *Bloomberg Business,* July 1.

Haidar, Jamal Ibrahim, and Takeo Hoshi. 2015. "Implementing Structural Reforms in Abenomics: How to Reduce the Cost of Doing Business in Japan." Working Paper no. 21507. Cambridge, Mass.: National Bureau of Economic Research.

Harada, Yutaka. 2014. *Reflationalist Economics That Saved Japan* [in Japanese]. Tokyo: Nikkei.

Hausman, Joshua K., and Johannes F. Wieland. 2014. "Abenomics: Preliminary Analysis and Outlook." *Brookings Papers on Economic Activity,* Spring: 1–63.

IMF (International Monetary Fund). 2014. "2014 Article IV Consultation—Staff Report; and Press Release." Country Report no. 14/236. Washington.

———. 2015a. *World Economic Outlook: Uneven Growth—Short- and Long-Term Factors,* April. Washington.

———. 2015b. "2015 Article IV Consultation—Press Release; Staff Report; and Statement by the Executive Director for Japan." Country Report no. 15/197. Washington.

Ito, Takatoshi. 2014. "We Are All QE-sians Now." Discussion Paper no. 2014-E-5. Bank of Japan, Institute for Monetary and Economic Studies.

Ito, Tatsuo, and Takashi Nakamichi. 2015. "Ex-Toyota Executive Tapped as New Bank of Japan Board Member." *Wall Street Journal,* April 21.

Jones, Charles I. 2015. "The Facts of Economic Growth." Working Paper no. 21142. Cambridge, Mass.: National Bureau of Economic Research.

Kawata, Hiroshi, and Saori Naganuma. 2010. "Labor Force Participation Rate in Japan." Review no. 2010-E-7. Bank of Japan.

Keen, Michael, Mahmood Pradhan, Kenneth Kang, and Ruud de Mooij. 2011. "Raising the Consumption Tax in Japan: Why, When, How?" Staff Discussion Note no. 11/13. Washington: International Monetary Fund.

Kimball, Miles. 2013. "Breaking Through the Zero Lower Bound." Paper prepared for Cognitive Economics Lecture Series. Coventry: University of Warwick. http://www2.warwick.ac.uk/fac/soc/economics/events/seminars-workshops-conferences/conferences/cognitiveeconomics/breaking_through_the_zero_lower_bound.pdf (updated 2015 version with Ruchir Agarwal available as IMF Working Paper no. 15/224)

Kitanaka, Anna, Shigeki Nozawa, and Yoshiaki Nohara. 2014. "Japan's Pension Fund Cutting Local Bonds to Buy Equities." *Bloomberg Business,* October 31.

Koo, Richard C. 2003. *Balance Sheet Recession: Japan's Struggle with Uncharted Economics and Its Global Implications.* Hoboken, N.J.: Wiley.

Krugman, Paul. 2013. "PPP and Japanese Inflation Expectations (Extremely Wonkish)." The Conscience of a Liberal (blog post). *New York Times,* October 27.

Kubota, Yoko. 2015a. "Nissan, Honda to Ship More Cars from Home." *Wall Street Journal,* July 10.

———. 2015b. "Toyota Posts Record Profit, Fueled by Weak Yen." *Wall Street Journal,* August 4.

Kumar, Saten, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko. 2015. "Inflation Targeting Does Not Anchor Inflation Expectations: Evidence from Firms in New Zealand." In present volume of *Brookings Papers on Economic Activity.*

Kuroda, Haruhiko. 2013. "Overcoming Deflation and After." Speech at the Meeting of Councilors of Nippon Keidanren (Japan Business Federation), Tokyo, December 25.

Kuttner, Kenneth N., and Adam S. Posen. 2001. "The Great Recession: Lessons for Macroeconomic Policy from Japan." *Brookings Papers on Economic Activity,* no. 2: 93–185.

Mankiw, N. Gregory, and Ricardo Reis. 2002. "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *Quarterly Journal of Economics* 117, no. 4: 1295–328.

Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers. 2004. "Disagreement about Inflation Expectations." *NBER Macroeconomics Annual* 18: 209–48.

McKay, Alisdair, Emi Nakamura, and Jón Steinsson. 2015. "The Power of Forward Guidance Revisited." Working Paper no. 20882. Cambridge, Mass.: National Bureau of Economic Research.

Nakamichi, Takashi, and Tatsuo Ito. 2015. "Abe Aims to Tip Balance on BOJ Board." *Wall Street Journal,* February 25.

Nakamichi, Takashi, and Megumi Fujikawa. 2015. "Japan Government Presses Companies to Raise Wages." *Wall Street Journal,* January 21.

Nakamichi, Takashi, and Yoko Kubota. 2015. "Japanese Companies Grant Biggest Pay Raises in Years." *Wall Street Journal,* March 18.

Patrick, Hugh. 2014. "Japan's Abenomics: So Far So Good, But Will It Prevail?" Occasional Paper no. 67. New York: Columbia Business School, Center on Japanese Economy and Business.

Posen, Adam S. 2014. "An American's Assessment of Abenomics at Mid-Term." In *Lessons from Decades Lost: Economic Challenges and Opportunities Facing Japan and the United States.* Briefing no. 14-4. Washington: Peterson Institute for International Economics.

Romer, Christina D. 1992. "What Ended the Great Depression?" *Journal of Economic History* 52, no. 4: 757–84.

———. 2014. "It Takes a Regime Shift: Recent Developments in Japanese Monetary Policy through the Lens of the Great Depression." *NBER Macroeconomics Annual* 28: 383–400.

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

Rose, Jonathan D. 2010. "Hoover's Truce: Wage Rigidity in the Onset of the Great Depression." *Journal of Economic History* 70, no. 4: 843–70.

Sommer, Martin. 2009. "Why Are Japanese Wages So Sluggish?" Working Paper no. 09/97. Washington: International Monetary Fund.

Svensson, Lars E.O. 2003. "Escaping from a Liquidity Trap and Deflation: The Foolproof Way and Others." *Journal of Economic Perspectives* 17, no. 4: 145–66.

Temin, Peter, and Barrie A. Wigmore. 1990. "The End of One Big Deflation." *Explorations in Economic History* 27, no. 4: 483–502.

Werning, Iván. 2015. "Incomplete Markets and Aggregate Demand." Working Paper no. 21448. Cambridge, Mass.: National Bureau of Economic Research.

Wieland, Johannes F., and Mu-Jeung Yang. 2015. "Financial Dampening." Working Paper (version 2.1). http://faculty.washington.edu/mjyang/research/findamp.pdf

Comments and Discussion

COMMENT BY

ADAM POSEN This paper by Joshua Hausman and Johannes Wieland is in a sense a direct reply to the paper Kenneth Kuttner and I wrote for *Brookings Papers* in 2001 (Kuttner and Posen 2001). It asks how we should understand the underperformance of Japan in recent years as opposed to during the so-called Lost Decade of roughly 1990 to 2002. Our assessment 14 years ago was that Japan's Lost Decade was largely understandable from standard textbook, if not undergraduate, macroeconomics. Excessive fiscal tightening, insufficient monetary stimulus, and a repeated failure to recapitalize the banking system in a timely manner all straightforwardly contributed to the lengthening of the initial post-bubble recession. The interesting question raised by Hausman and Wieland in this paper, and increasingly by others around the world, is this: Has recent underperformance in Japan become a bigger puzzle? In other words, since Abenomics has basically done many of the things that we were all calling for—including aggressive monetary stimulus with a forward-looking positive inflation target—why has it not worked better?

Hausman and Wieland give us a partial answer, and in particular their work on the consumption patterns of the Japanese populace in response both to monetary stimulus and to fiscal contraction is indeed new and provocative. Nonetheless, I feel that their paper misses two critical aspects. First, what was going on globally at the time? Is this mystery something idiosyncratic and specific to Japan, or a more general mystery all around the world? Second, is the mystery in some sense worse than they make it out to be, because they do not take into account the nature of the structural reforms that Abenomics has already put in place?

In essence, the standard trope about lost *decades* (plural), rather than one lost decade, remains exaggerated for Japan. As Hausman and Wieland,

among others, have pointed out, per capita GDP growth in Japan since 2002 has actually been quite good, especially in comparison to the rest of the G7 nations, even leaving aside the global financial crisis. Yet the questions about fiscal sustainability and sticky deflationary expectations are troubling. Because they fit into the broader international problem, arguably faced in the United States and western Europe as well, I would argue that the questions raised by Abenomics in Japan today are indeed even more troubling than the impression left by Hausman and Wieland would indicate. First, their discussion of monetary credibility misses the point of just how much the expectation-based arguments have failed in Japan, even when everything was in place for them to succeed. Second, given that these macroeconomic failures have happened simultaneously worldwide, the question remains as to what can be done to raise inflation and reduce public debt, whether in Japan or in other countries essentially sharing the same problem.

ASSESSMENT OF ABENOMICS TO DATE The basic fact that has to be reconciled is Japan's persistently low inflation and weak consumption since Prime Minister Shinzō Abe took office in January 2013 and launched a comprehensive reflation program with the cooperation of the Bank of Japan. Is this, as it seems, a failure of aggressive stimulus and coordinated macroeconomic policy—as I and many others have been advocating for Japan? Hausman and Wieland advance the discussion by focusing in on a very specific seeming contradiction: Two and a half years of openly declared forward-looking monetary stimulus have raised inflation above zero, but they have not led to any sustained upward movement of inflation beyond that. Meanwhile, the consumption tax hike of 2014 had a devastating and surprisingly persistent impact on household consumption. How weird is this?

If we go beyond academic theorizing about monetary stimulus through forward guidance, which I criticized at the time (Posen 2012), it is actually less surprising. Numerous market participants and policy observers expected that the pass-through to inflation would be limited from yen depreciation, even though announcements from Abe were seen as moving the yen before the Bank of Japan did anything. This reflected both Japanese history and the more recent experience of diminished exchange rate pass-through for the major economies at the upper end of the international division of labor. The United Kingdom's lack of sustained inflation pass-through following a 25 percent depreciation in 2008 is the most telling example—one that has now been followed by Japan, which saw an even bigger depreciation, from ¥79 to ¥120 to the dollar, without appreciable

pass-through to ongoing inflation. In a world where Japan has integrated supply chains throughout Asia and elsewhere, and given years of monetary restraint, it should not have surprised anyone that the pass-through would be limited. The recurring invocation of the "foolproof way" to inflation in Japan through depreciation (Svensson 2003) can be set aside as failed, as it would be difficult to imagine a more credible, large, and sustained depreciation than what has occurred.

Similarly, any surprise that a consumption tax hike would have a large short-term negative impact on consumption should not have been so great. Hausman and Wieland speak in their paper as though there was great uncertainty and debate about the size of multipliers for fiscal policy in Japan. Again, as with exchange rate pass-through, they paid too much attention to their academic colleagues and too little attention to actual observers and analysts from the official sector and the applied policy world. There was a relatively strong consensus on the size of the multipliers in Japan as seen through the 1990s, which held up to subsequent events and examination (Kuttner and Posen 2002; Romer 2012). These were very sizable, on the order of 1.5-plus, and likely to remain so given a relatively closed economy.

Nonetheless, the exercise that Hausman and Wieland perform, looking at the microdata in a cross-section of Japanese consumers, is a significant contribution. The fact that neither debtor/creditor status nor age category makes any statistically significant difference in consumers' minor reaction to the monetary stimulus or major reaction to fiscal tightening under Abenomics is a real surprise and worth knowing. That said, given the long-standing evidence on short-term views of Japanese fiscal policy by households and the stickiness of inflation expectations by all but foreign investors (Posen 1998; Kuttner and Posen 2001), this is not entirely shocking. It is rather the lack of variation across Japanese cohorts that is the mystery. It also is of interest to those of us watching the policy impact to see how persistent the shock from the value added tax turned out to be, lasting nearly four quarters in its visible effects. This is something that Hausman and Wieland could not address with their cross-sectional data set.

The problem with making too much out of these puzzles is that many of these aspects or tensions in the standard models are being replicated in the advanced economies around the world. This is in direct contrast to the mid- and late-1990s, when Japan was clearly on a different path and the difference could be attributed to idiosyncratic factors, primarily policy mistakes. Unlike the 1990s, many of the observed phenomena in Japan today—low wage growth, low exchange rate pass-through, no stable short-run Phillips

curve, low long-term interest rates despite fiscal weakening, low response of imports to depreciation, among others—are simultaneously global patterns. Few instances elsewhere are as extreme as we are seeing in Japan, however, and few have been subjected to such radical shifts in monetary regime exchange rate valuation or fiscal projections, and therefore are less clean tests. Nonetheless, all these sticky nominal variables are now being seen in the United States, in most of western Europe, and arguably even in China and South Korea. Thus, there is a more interesting and deeper puzzle here than the Japan-specific issues that Hausman and Wieland raise (one I make no pretension to answering, but hope this exchange will stimulate others to take on).

There is an issue that Hausman and Wieland should have addressed more deeply, however. This regards the significant changes in the Japanese labor market that have taken place since Abenomics began. The authors sensibly acknowledge the increase in part-time workers as a share of the total workforce, and the share of flexible hours rather than traditional full-time Japanese employment patterns. They have not done quite enough to grapple either with the scale of rapid change in Japanese labor markets or with what we have learned in recent years about the political economy of labor market reform. There is clear evidence from the experience of the European Union over the last 20 years, most notably the Hartz IV reform in Germany in 2003, that labor supply reforms are deflationary in the short run. In fact, it often seems that labor supply reforms only bear fruit whenever the next strong recovery takes place, and not before.

In this regard, the paper also needs more discussion of Japan's demographics. While the country is aging—a fact directly picked up in their cross-sectional data set of consumption—there could be other dynamics at work as a result of demographics, including later-in-life rises in risk aversion among workers and more off-the-tax-books informal employment than is currently measured. But in the end, the authors' discussion of labor dynamics, and more broadly of Japanese performance, misses the fact that productivity performance has not been terrible. If one thinks about per capita growth as reported in the paper and realizes that all of it has to have come from productivity growth, given that there has been little capital deepening in Japan in the last decade, it has not been a bad performance.

In this regard, readers of the paper should be reminded that the record of Abenomics's so-called third arrow, the implementation that brought structural reforms, is far better than implied by this paper (though to be fair, the authors are not trying to do an overall assessment of these issues). For the

record, there have been two and a half major structural reforms undertaken by the Abe government to date:

—"Womenomics," as the policy of incentives to bring more women into the workforce has been dubbed, has raised female labor force participation by over 2 percent in two and a half years. This means that more than 750,000 women have rejoined a workforce of 63.5 million. Many of these women have joined the workforce on a part-time or flexible basis, but we know from the Nordic experience that this is part of the way one retains women in the workforce, so it should not be discounted. This clearly has had a disinflationary if not deflationary effect on wages, not just by increasing supply but also by bringing in people who—both on seniority grounds and sexist grounds—are paid less than comparable male workers. Nonetheless, it is a major structural reform with huge long-term implications.

—Corporate governance in Japan has been improving, though of course it has not been markedly transformed. In particular, the transparency of corporate accounting and accountability for those accounts in Japan has clearly risen. The scandals at Olympus and Toshiba[1] show how, at least for publicly listed companies in Japan, previously accepted behaviors are no longer acceptable. The larger inflows of foreign investment into Japanese equity baskets and directly into companies are both cause for and effect of these reforms. There is actually a positive cycle in this area: Some of the rise in equity values in Japan is arguably attributable to improved corporate governance and transparency, not just to yen depreciation, unlocking values that have been hidden. Undoubtedly, the overall rise of profitability in Japanese multinational companies is largely due to export demand and to the country's recovery from deflation, but the structural reforms have mattered. We are indeed seeing an increase in two-way flows of foreign direct investment inward and outward from Japan, which is consistent with the story.

—Agriculture is also undergoing significant reform, though so far it is only partial. This reform is likely to accelerate and to then be made permanent with the Trans-Pacific Partnership (TPP) trade agreement. The Abe government has taken on Japan Agricultural Cooperatives (JA), which is like a hybrid of the National Rifle Association interest group in the United States and the European Union's Common Agricultural Policy. In fact, rather than waiting for the average Japanese farmers to die out and have their farms be consolidated—which would probably be in the next 5 to

1. See Russell (2015).

7 years—the Abe government has directly moved to increase efficient scale and competitiveness in agriculture. Price supports and tariffs are already being broken down for several major products, including previously untouchable goods such as dairy and pork. Should the TPP agreement as negotiated pass in the coming year, there will be further rapid change in Japanese agriculture, given the commitments that the government has made.

Hausman and Wieland rightly give considerable attention to whether long stretches of underperformance in GDP and productivity growth lead to self-fulfilling downward projections of potential output in Japan (an issue I raised in Posen [1998]). This set of reforms raises an intriguing possibility, namely that for at least a few years, potential growth might be raised on both labor supply and productivity grounds.

DOES ABENOMICS SUPPORT OR DISCREDIT THE MACROECONOMIC MAINSTREAM OF TODAY? As in many other places where it was put to the test in recent years by the global financial crisis, the forward-looking expectations and credibility-based view of monetary policy comes off very poorly with respect to explaining Abenomics's outcomes. Some warned 15 years ago that it would not be enough simply to promise irresponsible policy or future higher inflation—Japan, they argued, would need real growth (beyond closing the output gap or achieving the NAIRU[2] level of unemployment) to get inflation up. Putting it differently, the combination of the visible regime change at the Bank of Japan with a forward-looking 2 percent inflation target promise, strong backing from the government (some would say even with erosion of central bank independence), and a sustained sizable devaluation in the yen should have been enough to raise inflation expectations—if indeed inflation expectations were ever to be malleable to such "credible commitments." One can always give the excuse that this multifaceted public effort with political backing at the highest level was not sufficiently credible, but then one ends up sounding like Margaret Thatcher trying to defend monetary targeting in the 1980s despite its repeated failures. If Abenomics and the Bank of Japan program under Governor Haruhiko Kuroda were not a credible precommitment to raise inflation, it is very difficult to imagine what would have constituted one.[3]

2. NAIRU stands for non-accelerating inflation rate of unemployment.

3. Juxtaposing this with Kumar and others' (2015) study of inflation anchoring in New Zealand—the other paper in this BPEA volume on inflation targeting—one finds that the hypothesis that expectations are the main channel for policy transmission is failing there as well, despite an a priori credible monetary regime with clear, transparent inflation targets that track over many years.

This may be a little unfair—more to the Bank of Japan than to the theory—in that the combination of labor market changes (notably the addition of many part-time female workers) and various global disinflationary forces (such as the sustained drop in energy prices) can account for some of the shortfalls in the Bank of Japan's progress towards its 2 percent inflation target. Based on the latest data, core-core inflation[4] is trending up and is at 0.9 percent on a year-over-year basis, and the Bank of Japan board members forecast that the 2 percent target will be reached in the second half of 2017.

But the hidden surprise, unnoted in the paper, is that there are no balance sheet effects present to damp down the inflation response now, unlike in the early 1990s and the 2000s. Again, this is not dissimilar from, say, the situation in the United States and some western European countries, but it is more extreme in Japan's case; since the mid-2000s, corporate balance sheets in Japan have been extremely strong, and household balance sheets have been in good shape as well. In addition, unlike in the United States and western Europe, and unlike in Japan in the 1990s, today the banking system in Japan is well capitalized and in demonstrably good shape, having ridden out the financial crisis with little problem. If anything, the puzzle becomes greater because these strong balance sheets, at a minimum, should have partially offset whatever downward pressure wages and energy exerted on inflation over the short run.

Fiscal policy comes closer to being what was expected in standard macro theory, or rather, standard policy and applied macro theory (not the fanciful New Keynesian models, let alone real business cycle models where fiscal policy is ineffective). As noted, the multipliers on fiscal policy in Japan are large but not unexpectedly so; it is the persistence of the value added tax shock that was a surprise, especially given the ample forewarning of its coming and its being embedded in a putative series of tax hikes. The underlying challenge is in the (rational) expectations channel. It has been decades since economists took Ricardian equivalence literally, but some notion of forward-looking markets, and even average citizens' awareness, with respect to Japan's well-documented and unbending rise in public debt, should have been expected. Yet we see little sign of such behavior, from household savings, which show little response to movements in fiscal policy, to long bond rates, which also remain largely unperturbed, even as deficits mount. The outcome of the authors' interesting exercise, showing the uniformity of consumption impact across age and even creditor

4. Japan's "core-core" inflation index excludes food and energy prices, and is similar to the core inflation index used in the United States.

status, is therefore somewhat troubling. Here at least there does seem to be a meaningful difference between Japan and other countries, with cross-sectional differences of age and financial status seeming to matter more in Japan for fiscal policy response. This bears further examination.

But in a world where some still talk about fiscal theories of the price level and treatment of debt sustainability as immediate issues, the stability of Japanese savings and interest rates must be seen as a major challenge. It also is a challenge to the idea of simple distinctions between permanent and temporary tax policy impacts, since we are getting to the end of the fiscal road in Japan, and the whole point of the consumption tax increase was that it was still only one in a precommitted series. This is not to say that Japan's even net debt (currently at 160 percent of GDP) is in any sense on a sustainable path, given demographics and health care commitments. It is to say that some measures of financial repression and monetary financing are clearly more effective at calming individual behavior than one would have been led to believe by much of the economics profession in recent years. Meanwhile, the markets are littered with the tombstones of hedge funds that dared to bet on expectations of a Japanese fiscal collapse. This will no doubt change should inflation be sustained for a long period and long bond interest rates eventually begin to rise—at that point, the interest payment dynamics will rapidly crowd out all other activities in the Japanese budget and demand a response. But for a surprisingly long meantime, forward-looking behavior on the part of Japanese households, and even Japanese government bond markets, seems to be absent.

I would like to emphasize that there are still three surprises in the underperformance of Abenomics, even though I give the policy package a more sympathetic reading on structural reform and even monetary policy impact than Hausman and Wieland do. The first surprise, as I mentioned earlier, is that clean balance sheets across the Japanese economy seem to have brought less benefit in terms of growth and investment than was expected. There is no question that the resolution of the banking crisis in 2003 undertaken by Heizō Takenaka, then minister of financial services, which included recapitalization and consolidation of the banking sector, was a necessary condition to get Japan out of its worst Lost Decade. But it must be reckoned as an asymmetry to deal with, one similar to what we are seeing in the United States at present, that while fixing a banking crisis prevents bad outcomes, it does not seem to stimulate good outcomes.

A second surprise from Abenomics's underperformance leads me to ask whether economists should just stop talking about credibility of macroeconomic policy altogether. If forward-looking behavior matters, we

should have seen some great response to the uncertainty about the path of future consumption tax hikes induced by the Abe government's decision not to proceed with the next scheduled one (which was to have been in spring 2015). It is hard to get away from this fact. Similarly, if credibility was key, the Bank of Japan's inflation commitment should have been sufficient to raise inflation to target on trend.

The third surprise is that while there have been meaningful steps, albeit perhaps insufficient progress in both labor market reform and corporate governance reform in Japan, as reform is defined by the Organization for Economic Cooperation and Development or according to western textbooks, there has been little obvious benefit to productivity or to reallocation of resources. These structural reforms therefore must be thought of as possibly overrated, given the lack of support beyond the equity market, and the certainly disappointing results with respect to real GDP growth. Again, one can hope that this is attributable to these kinds of structural reforms, particularly labor market reforms, being disinflationary in the short run, as is evident in southern Europe. This too bears further scrutiny.

WHAT CHALLENGES DOES ABENOMICS PRESENT TO MACROECONOMIC POLICY?

Remember, the message that came out of past research was that in Japan from 1990 to 2003, or arguably from 1985 to 2007, macroeconomic policy worked as expected. When monetary policy stopped being deflationary and started to get ahead of the curve, many things improved. When fiscal policy was tightened or loosened, despite the overhang of private and public debt, large consumption responses were seen. The restoration of bank capital made a major difference to the behavior of the economy, although as mentioned earlier it ended the downside risk more than it raised the upside performance. Has something changed to make this less the case in Japan now, and is this something one should look at in other countries as well?

The usual catchall these days for explaining Japanese exceptionalism is demographics, to say that fundamentally as the population gets older, various perverse behaviors start to dominate, including extreme risk aversion and underinvestment. Yet these channels have not been well specified, and it is clear from cross-sectional growth regressions that the declining population is usually good for per capita growth, so things cannot be quite that simple. As many note, unfavorable demographics with respect to the working-age population are hardly a problem unique to Japan. Additionally, Japan is now a more open economy and more market-oriented, if not more subject to market discipline on any a priori observable grounds than it was in 2003 or 1993. Put simply, this should go the other way, and make

Japan's response to macroeconomic policy more like it is in other countries (except for fiscal leakage, which is a second-order issue), and more like what the textbooks describe. That is not what has happened, though.

Before economists get caught up in the overly cutesy and complex discussions of behavioral economics, which have replaced the previous generation's technical fascination with real business cycle models, they have to look at much more simplistic (and not just simple) microfoundations. In short, the rational expectations revolution that permeated all our macroeconomic models for the last 30-plus years is probably a grievous mistake if Japan is anything to go by—or if the global financial crisis counts, for that matter. Of course, it pains everyone to say that, and it would require a huge rethink with no obvious ready-made alternative.

But if there is one message from Japan's example as a macro puzzle, it is the near total absence of forward-looking behavior, particularly in response to monetary policy, but also to fiscal problems. This is evident even among Japanese businessmen and investors, let alone being widespread among Japanese households. One must confront the idea that there is some real stickiness in expectations and even in regimes, at least once one has entered a near-deflation low-growth environment. In a Brazil or an India today, let alone in smaller open economies, one still legitimately watches for what the 1970s taught us to fear: inflation spirals, vertical Phillips curves, and fiscal dominance, which all certainly still have relevance. But those patterns might not be relevant for the low-inflation world in which Japan and other advanced economies now find themselves.

The policy research challenge is to further examine the global forces that lie behind the current persistently low inflation levels, and in fact lie behind the parallel changes in wage share, consumption trends, and investment appetite across the advanced world. Are these outcomes the result of direct spillovers, or of common policy approaches (and mistakes), or of some underlying transnational forces at work? Both real business cycle models and their hybrid children in New Keynesian open economy models have done a poor job of fitting what happened during the crisis—but more importantly, and even more confusingly, they have failed to predict or explain what has happened *since* the crisis. It is in this sense that the evidence presented in Hausman and Wieland for Japan's anemic recovery despite monetary stimulus is truly troubling, because it is being echoed in the euro area and in the United States at present.

One can talk about a common downshift in productivity growth, which certainly is seen in the data, but all else equal, that should not keep market economies with well-capitalized banking systems from having a positive

response to monetary stimulus, whatever the limit on potential. Remember, it used to be considered the main policy problem that monetary policy had a proclivity to successfully and persistently stimulate economies well above potential. Moreover, as Japan illustrates, along with the United States and arguably much of western Europe, the productivity slowdown was started well before the global financial crisis. Using a productivity trend break to explain more than the immediate precrisis buildup and postcrisis bust therefore seems to be a stretch.

Japan's experience, as pointed to by Hausman and Wieland, also raises significant issues having to do with exchange rate pass-through and the trade balance. We spend a lot of time in Washington and other capitals, as well as in central banks around the world, hoping that the exchange rate proves to be a major mode for transmitting monetary policy in the real economy—and fearing that some other country will export their unemployment. People who dispute my very negative view on the expectations channel of monetary policy will point to the substantial depreciation of the yen since Abe won election as prime minister in December 2012 as evidence that the expectations channel was working. Yet what we have seen in country after country, and most strikingly in Japan and the United Kingdom, is a shortfall on past benchmarks of net export response to large sustained depreciations. In fact, the response seems to be even more diminished on the import side than the export side. This is troubling, because on the export side it is easier to rationalize companies taking profits in their home currency and maintaining market share, as there is good indication Japanese multinationals have done in this cycle. There is reason to think about a diminished net export effect for countries whose trade heavily involves industry and particularly, intrafirm supply chains, as is certainly the case for Japan's network throughout Asia. Nonetheless, that integration of production does not seem to fully explain the limited net export improvement, especially in societies where household consumption of imported goods remains high and where the manufacturing share of the GDP is declining.

Bringing this back to the original questions from Hausman and Wieland's focus on monetary policy, one can summarize the puzzle this way: How could such a large exchange rate depreciation, seemingly caused by a significant monetary regime shift with commitment, have had so little effect on general inflation at home? Of course, it was a matter of deep faith that inflation-targeting regimes' "well-anchored expectations" would allow exchange rate shifts to be treated as first-round impacts and not passed through. The trouble with that is that the whole point of the Abenomics

exercise in Japan was to explicitly reanchor deflation expectations upward, from an unclear commitment to a positive inflation target to a clear one. Why that should be less credible than keeping inflation expectations anchored is difficult to understand, especially since Japan was coming out of a deflationary period acknowledged to be harmful.

It is even more puzzling given that all the forward-looking models from Finn Kydland, Edward Prescott, Robert Barro, and David Gordon onward assume that there is always doubt about the credibility of the central banks' commitment to stay anti-inflation. In other words, in the mainstream macro policy literature, there is a fundamental assumed asymmetry in monetary policy that makes it easier to bring inflation up than to take it down. Again, this clearly seems to be disproved by Japan's experience, as well as by recent though less stark experiences in Europe and the United States. There are some very serious questions being provoked here by Hausman and Wieland, although in their paper's section on central bank credibility they indulge too much in trying to reconcile these anomalies with the existing literature. Economists will have to think much harder about what Japan means this time than they did 15 years ago, for the message now is much more disruptive to the standard macroeconomic understanding.

REFERENCES FOR THE POSEN COMMENT

Kumar, Saten, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko. 2015. "Inflation Targeting Does Not Anchor Inflation Expectations: Evidence from Firms in New Zealand." In present volume of *Brookings Papers on Economic Activity.*

Kuttner, Kenneth N., and Adam S. Posen. 2001. "The Great Recession: Lessons for Macroeconomic Policy from Japan." *Brookings Papers on Economic Activity,* no. 2: 93–185.

———. 2002. "Fiscal Policy Effectiveness in Japan." *Journal of the Japanese and International Economies* 16, no. 4: 536–58.

Posen, Adam S. 1998. *Restoring Japan's Economic Growth.* Washington: Institute for International Economics.

———. 2012. "Commentary: Methods of Policy Accommodation at the Interest-Rate Lower Bound." In *Economic Policy Symposium Proceedings: The Changing Policy Landscape.* Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

Romer, David. 2012. "What Have We Learned about Fiscal Policy from the Crisis?" In *In the Wake of the Crisis: Leading Economists Reassess Economic Policy,* edited by Olivier Blanchard, David Romer, Michael Spence, and Joseph E. Stiglitz. MIT Press.

Russell, George W. 2015. "Will Toshiba's Scandal Bring About the Change Needed in Corporate Governance?" *GAA Accounting,* May 10. http://www.gaaaccounting.com/toshiba-scandal-corporate-governance

Svensson, Lars E. O. 2003. "Escaping from a Liquidity Trap and Deflation: The Foolproof Way and Others." *Journal of Economic Perspectives* 17, no. 4: 145–66.

GENERAL DISCUSSION Lars Svensson found the paper by Joshua Hausman and Johannes Wieland to be a very good update on their earlier research. In his own view, a bolder policy, one that did not rely mainly on just the pass-through from the exchange rates, might arguably have moved expectations in Japan more successfully than the current policy has done, thus increasing inflation expectations and reducing real interest rates. Such a policy, the "foolproof way" of escaping from a liquidity trap,[1] which he had advocated for Japan in 2000 and is somewhat similar to what the Czech National Bank is doing currently, includes three elements: (i) a currency depreciation of some 10 to 15 percent, with a corresponding exchange rate floor; (ii) a price-level target some 10 to 15 percent above the current price level; and (iii) an exit strategy, according to which the currency is floated and policy is returned to normal when the price-level target has been reached.

Svensson suggested that the outcome would have been much better if Japanese Prime Minister Shinzō Abe and Bank of Japan Governor Haruhiko Kuroda had been appointed 15 years ago and deployed their policy at that time. Instead, what has transpired in Japan is the entrenchment of a "deflationary mind-set," something the Bank of Japan board members have mentioned several times. Expectations among the Japanese are deeply entrenched at this point, and getting them out of that mind-set after 15 years is difficult.

Frederic Mishkin elaborated on the issues that discussant Adam Posen had raised regarding how demoralizing the outcomes from Japanese monetary policy have been. He had felt more strongly than Posen that expectations were very important and that managing expectations is a key element in good monetary policy. He and his colleagues expected much stronger effects in Japan from the expansion of its monetary policy. Japan's outcome

1. Lars E. O. Svensson, "The Zero Bound in an Open Economy: A Foolproof Way of Escaping from a Liquidity Trap," *Monetary and Economic Studies* 19, no. S-1 (2001): 277–312.

might demonstrate that raising inflation expectations is much more difficult than lowering them, and moreover this might be true globally.

Acknowledging that he is known to be a big proponent of inflation targeting, Mishkin said that when the focus is on how to keep inflation expectations down, it has worked well. But he and others have found it much more difficult to raise expectations, particularly during a long period of deflation.

Brad DeLong seconded Mishkin's comment, adding that the macroeconomic situation in Japan has not developed to Japan's advantage, even though economists had strong reasons to think the expectations channel was present based on historical examples. Franklin Roosevelt's New Deal and Neville Chamberlain's announcement that he was going to restore Great Britain's price level to its pre-Depression state both demonstrated the power of the expectations channel, and indeed the same happened when Japanese Finance Minister Takahashi Korekiyo announced his decision to go for reflation in Japan in the 1930s. It is a great puzzle that this time around it has not been working.

Robert Gordon wanted to reiterate something Posen had stated earlier about the paper in the present volume by Saten Kumar, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko about New Zealand, which showed that analyzing what professional forecasters think is not nearly sufficient because it may have very little to do with what the average price setter thinks. The paper by Hausman and Wieland pays too much attention to how expectations are formed and to central bank credibility, in Gordon's opinion, whereas if the New Zealand example carries over to Japan, the actors who are actually setting prices and wages might not even know what the central bank is. In sum, rather than looking at how the central bank affects expectations, he thought one should be studying how actual prices are set.

Martin Feldstein mentioned a proposal he had made several years earlier for increasing the expected inflation in Japan. It would be a balanced-budget tax change in which the government announces it is going to raise the value added tax by some certain amount—say 1 or 2 percent per year—and that it is going to balance that by cutting the personal income tax. This approach would avoid a Keynesian aggregate demand effect, but it would stimulate an expectation that prices were going to rise, which should increase aggregate spending. Feldstein continues to think this is a good idea.

Christopher Carroll complained about the fact that while macroeconomists have been open to the idea of developing microfoundations for the impact of labor supply and responses to tax policies, it has been

nearly taboo for them to research how people actually form expectations, including inflation expectations.

David Romer wished to highlight an aspect of the paper that Hausman had no time to discuss in the presentation, namely that when examining Japan's economy it is incredibly important to consider that the working-age population is falling. The headline output numbers may look dismal, but output per working-age population actually rose 3.2 percent in 2013 and 1.4 percent in 2014, both quite respectable numbers. This led him to wonder—though it seemed hard to believe—whether it was possible that forecasters in Japan had failed to understand that the working-age population was falling. Their 2007 forecast of GDP was for steady 2 percent growth up through today and beyond, and at least now the working-age population is falling at 1.5 percent a year, so this means that in 2007 they were actually forecasting roughly 3.5 percent growth in output per worker indefinitely. This suggests that in comparing past forecasts with actual outcomes and current forecasts, it may be important to consider the possibility that the past forecasts did not properly account for demographics, and that forecasters have only gradually incorporated the falling working-age population into their forecasts.

Posen concurred with Romer on this last point, adding that he found the forecasts especially strange since the Japanese government, the International Monetary Fund (IMF), the Organization for Economic Cooperation and Development (OECD), and the banks that issue forecasts all state that potential growth in Japan is well below 1 percent, possibly as low as 0.5 percent. The fact that forecasts continue to be published that forever project 2 percent growth struck Posen as weird.

Katharine Abraham added to the comments others had made about the importance of structural reforms to the long-run success of the Japanese economy. She underlined the striking fact that between now and 2050, projections are for the Japanese population to fall by a third, with a substantially higher fraction of that smaller population consisting of older people who traditionally have not been active in the labor force. In Abraham's view, structural reforms that could increase the size of the labor force should be central to the ongoing policy discussion in Japan.

She agreed with Posen's earlier comment that Japan has made progress in drawing women into the labor force. At the same time, much more needs to be done to make it possible for women to work at jobs that make full use of their capabilities. Historically there have been many disincentives for second earners in Japanese households to work more than minimal hours at low wages. In addition, there is the challenge of enabling

women to enter the labor force without depressing fertility even further. The tradition of men working long hours and not coming home until the small hours of the morning is incompatible with women working full-time and also having children. In short, fully integrating women into the labor force is going to require fundamental changes in the way work is organized in Japan.

Immigration is another policy area where, Abraham thought, any country in Japan's situation would be seriously engaged. In the United States, half of the net growth in population stems from immigration, but immigration to Japan is minimal. Her sense of the Japanese perspective on immigration, based on conversations with Japanese officials during a visit over the summer, is that they are not thinking seriously about this pathway. One Abe minister told her he was hopeful about making progress on immigration because the number of foreign tourists had risen a lot, enabling more Japanese people to meet foreigners and thereby become more open to the idea of immigrant labor. That exchange led her to conclude that Japan has a long way to go.

Justin Wolfers took exception to the pessimism about Abenomics he was hearing in the discussion, and it puzzled him because he felt an equally strong case could be made for optimism. For example, the paper itself showed that inflation expectations have risen in Japan, indicating that talking to the public about inflation does work. The lesson seemed to be that when one promises 2 percent inflation, one gets 1 percent inflation, so perhaps the rule should be that to get 2 percent inflation one should promise 4 percent. The second lesson seemed to be that when a good policy shock is followed by a bad global shock, the net effect is that the economy just muddles along, and in fact that is how things look in Japan. Perhaps more explicit counterfactual thinking is needed to separate out the effects, since the world was a different place when Abe was elected, and clearly it went on to become a worse place.

Wolfers's third point was that the authors' initial assessment, in their first paper on Abenomics,[2] had been pessimistic, even though stocks had already risen 62 percent. The authors dismissed the stock rise as over-optimistic betting on future corporate earnings. In the year since then, stocks rose another 36 percent. Clearly, *something* has been raising the net present value of future corporate earnings. That demonstrated optimism in

2. Joshua K. Hausman and Johannes F. Wieland, "Abenomics: Preliminary Analysis and Outlook," *Brookings Papers on Economic Activity*, Spring 2014: 1–63.

the market was more convincing to him than the modeling of professional economists. It could be due to a redistribution from labor to capital, from small firms to big, that is, from newcomers to existing firms. In any case, it would be worth knowing whether the authors thought any of the structural reforms had changed how one should interpret stock prices.

Seeking to offer a slightly different perspective on the puzzle of monetary policy, Ricardo Reis found it striking that Abenomics—which the authors describe as expanding monetary policy, expansion through fiscal policy, and structural reforms—signified an elected leader undermining the independence of the Bank of Japan. The failure to raise inflation expectations could also be seen as a failure of the time-consistency view of inflation, which is that if one takes away the central bank's independence one is left with a government that is pushing for inflation. One would expect higher inflation, which after all is the lesson from the work of Robert Barro, David Gordon, and Finn Kydland.

With that in mind, Reis wondered whether the authors could speak about how the Bank of Japan in the last two years reasserted its independence, or did not, and whether one could look through that prism to understand the evolution of inflation. It seemed to him that that ought to be a vital part of the agenda, and that the psychology literature offered plenty of tools to examine it with.

Wieland replied to the discussion. First, he wanted to underline that by no means did he and Hausman regard Japan's situation as a disaster. They agreed with Romer that once one adjusts for demographic changes, GDP growth in Japan looks reasonable. It still lags behind the United States, but its productivity growth over the last 20 years has only been about half a percentage point below the U.S. rate. What is disappointing is the gap between the current level of Japanese GDP and the pre–Great Recession demographically adjusted trend.

In response to Wolfers's comment about raising inflation expectations further, he wondered by how much they would need to be raised. Would doubling the expected rate, as Wolfers suggested, be sufficient? His own optimistic interpretation was that the policies have at least been moving things in the right direction, since inflation expectations in Japan have gone up, and there is still scope for raising expectations higher. Nevertheless, he and Hausman hesitated to commit to a specific value for a higher inflation target. As Posen argued, it remains unclear what the right model for Japan is, so simply extrapolating and recommending that the government double the expected rate in order to achieve twice the effect is not justifiable.

He added that as found in the paper on New Zealand by Kumar and others, in Japan there has been a disconnect between inflation facts and the public's inflation perceptions, with opinion surveys by the Bank of Japan finding that consumers thought inflation had been averaging 4 percent for the previous 10 years and expected it to continue at 4 percent per year.

The increase in the labor force participation of women has been a great success for Abenomics, in Wieland's view. Nevertheless, when he and Hausman tried to determine what policies had been enacted to cause this change, they could only find a small set of policies. They remain uncertain how to square the outsized effects with the actual policies implemented.

Commenting on Reis's discussion of the time-inconsistency literature, he noted that some appointments to the Bank of Japan's governing board were of people who were thought to be friendly to the Abenomics agenda, and yet those appointments led to no obvious movements in inflation expectations. While one does not know the counterfactual, one did not see outsized movements in inflation expectations stemming from political interference in the central bank, which may speak to the theory of time inconsistency.

To Feldstein's suggestion of influencing consumer behavior through a value added tax, Wieland responded that Japanese corporations have been very resistant to continually changing prices in this way.

RICARDO REIS
London School of Economics

Looking for a Success in the Euro Crisis Adjustment Programs: The Case of Portugal

ABSTRACT Portugal's economic adjustment program in 2010–14 under the troika was extensive, designed to address both its large debt and its anemic growth, so it might serve as a blueprint for reforms in the eurozone. This paper argues that, based on a diagnosis of the underlying problems of the Portuguese economy, the adjustment program failed to definitively address the public finance problems but succeeded in opening a pathway for reforms in the economy. On the negative side, public debt is still high, primary surpluses improved only modestly, and public spending barely fell as the problem of ever-rising pension payments remained unsolved. On the positive side, unemployment fell sharply, exports and the current account balance rose, capital and labor reallocated to more productive and tradable sectors, and the economy is growing faster than the European Union average for the first time in 15 years.

From the start, the euro crisis posed a unique challenge to crisis management. In many ways, the events of 2010–11 in Greece, Ireland, Portugal, and Spain resembled a classic sudden stop. But dealing with the ensuing recession was bound to be difficult, as there were no European institutions set up to deal with a crisis of this type and magnitude. There was no currency to devalue, no independent central bank to back up and resolve struggling national banks, and few private bondholders to arrange for a debt write-down. Large and legally protected welfare states are everywhere hard to reform, and these European countries are no exception. As a result, the adjustment programs for these four countries were partly improvised and unique in their features, and their effectiveness was in question from the start.

Still, if the euro survives, there will surely be new crises in the future. In turn, as other regions in the world choose different forms of economic integration, they would like to learn what mistakes to avoid in following

the European example. Inspecting the adjustment programs put in place during the past few years is therefore in order.

Looking for a failed adjustment program is easy: no matter where the blame lies, it is undeniable that the Greek program has failed, with grim consequences for its population. Pointing to a successful adjustment program that could serve as a fair counterexample is more difficult. Ireland, for example, had started many of the reforms to its banking sector and public finances before its adjustment program began, and there were few macroeconomic measures in that program. Spain received financial assistance to recapitalize its banks with conditions on implementing reforms in its financial sector, but it did not enter a full-fledged macroeconomic adjustment program at the hands of the International Monetary Fund (IMF). This leaves Portugal as a potential example of success to counterbalance the failure in Greece.

Portugal is a good case to focus on for many reasons. Its adjustment program is already complete, and it consisted of an exhaustive list of reforms almost all of which were fully implemented. Coming after the interventions in Greece and Ireland, Portugal's program benefited from the accumulation of some experience. Finally, while all four countries have their idiosyncrasies, Portugal's crisis did not involve a housing price boom, nor extreme fiscal profligacy, but was mostly due to a complete lack of productivity and economic growth since 2000. Success in reversing that slump might offer lessons on how to raise the disappointing prospects for economic growth in the euro area as a whole.

There are two public views on the success of the adjustment program. One is captured by the statement of the influential German finance minister Wolfgang Schäuble, commenting in June 2014 on Portugal's announcement of the end of its program with the IMF:

> Portugal's reform efforts have paid off. Today's decision by the government in Lisbon is proof of this. Portugal no longer needs European assistance and can stand on its own two feet again. This is a major success. Capital market confidence has returned, and rightly so. (German Federal Ministry of Finance 2014)

From a narrow perspective, with success defined as being able to resume sovereign borrowing, Portugal delivered. The Portuguese state became able to borrow again and at moderate 10-year interest rates, both at the end of the program (at 3.5 percent) and thereafter. Throughout 2015, average monthly 10-year interest rates never exceeded 3 percent, despite the Greek crisis.[1]

1. The sources for the data mentioned in the text are varied; they are described in the online appendix. Online appendixes for all papers in this volume may be found on the *Brookings Papers* web page, www.brookings.edu/bpea, under "Past Editions."

Equally important, the troika[2] extended the maturity of the Portuguese official debt and reduced interest payments, and the Portuguese debt office successfully extended the maturity of the outstanding debt, with 10-year issuances throughout 2014 and 2015. As a result, the average maturity of the debt increased from 6 years in 2010 to more than 8 years at the start of 2015, reducing rollover risk (Reis 2015). Another debt crisis is unlikely in the near future.

A different view from Schäuble's was expressed one year later by Paul Krugman (2015) in an editorial that included Portugal among "Europe's Many Economic Disasters," where he stated:

> Portugal has also obediently implemented harsh austerity—and is 6 percent poorer than it used to be.

From the perspective of macroeconomic performance, the program seems to be a failure, with real GDP per capita 4.9 percent lower in 2014 than it was in 2010, and total employment falling from 4.9 million to 4.5 million. If success is judged as a rebound of the economy from its prolonged depression, then there is little to celebrate.

There is a simple way to reconcile these two opposing views. The first view focuses on public finances, where the program has delivered, while the second view argues that its consequences were a macroeconomic disaster. Both views could be right, with success in stabilizing public finances accompanied by few gains in getting the economy out of its slump. This paper argues, however, that both views are most likely wrong. Its verdict on the adjustment program is actually the opposite of the two views represented above: there are promising changes in the structure of the economy, but public finances remain far from a path that lowers the public debt.

It is hard to judge the success of a program without knowing what its criteria are and what counterfactual one is using to judge it against. My approach is to look at the progress made in solving Portugal's underlying structural problems and in addressing the four key challenges that Portugal faced at the height of the crisis: paying for large past debts, controlling future public spending, restarting economic growth and lowering unemployment, and improving competitiveness and capital allocation. I begin in section I by providing a diagnosis of Portugal's slump and crash. In sections II through V, I then measure success conditionally on the economy's

2. The "troika" is the common moniker for the trio of international institutions that have dominated financial rescue operations in Europe since the financial crisis: the European Commission, the European Central Bank, and the International Monetary Fund.

diseases to determine whether the program helped to heal them.[3] Another approach would have been to compare the adjustment program to what would have been ideal, that is, if the best policies had been followed. While there have definitely been many mistakes, I leave for others the job of highlighting them and arguing whether they are only clear now with the benefit of hindsight.

I. Diagnosis of the Crisis

Portugal requested international help in April 2011 and officially agreed to terms one month later. This came after a run-up in 10-year interest rates on government bonds, which reached 9.6 percent in May, up from 5.0 percent one year earlier. The government had difficulty rolling over bonds that were coming due and signed a 3-year agreement with the troika to secure financing of up to €78 billion, which expired on June 30, 2014.

The euro crisis arose when large capital flows from the core to the periphery of Europe, which had built up since the introduction of the euro, suddenly reversed in 2009–10. Without a currency to depreciate between different regions of the eurozone, the large and sudden contraction in the current account deficit required a large contraction in domestic consumption and investment, driving these economies into recession. A fall in the real exchange rate was required, but the usual rigidities that slow the adjustment of prices and wages led to a large and prolonged increase in unemployment. This is the traditional side of the crisis (Shambaugh 2012; Blanchard 2013).

New to this sudden stop, the capital flows across borders were intermediated by banks and largely funded through the interbank market (Brunnermeier and Reis 2015). In the European periphery, banks and capital markets lacked the depth to allocate the large inflows that came with financial integration, likely misallocating them into unproductive nontradable sectors. A flight to safety in response to higher risk aversion following the 2008 financial crisis had a cross-border dimension in Europe. The sudden stop came with fire sales in financial markets and falls in bank capital that led to large contractions in domestic credit.

Another novel and unique feature to the euro crisis is what has been labeled the "diabolic loop" or the "doom loop" between banks and sovereigns (Brunnermeier and others 2011; Obstfeld 2013). European banks

3. The European Commission (2014) and Jorge (2014) provide alternative evaluations, more favorable and more critical, respectively.

held large amounts of sovereign bonds. As economic activity slowed and public deficits rose, fears about sovereign default led to falls in the prices of government bonds, large losses in banks' holdings, and further fire sales and contractions in credit, deepening the recession. Once the crisis was in motion, the diabolic loop worsened because banks would offset the sudden stop of private capital by pledging government bonds as collateral at the European Central Bank (ECB) to obtain public financing. Together with the official troika bailout programs, this implied that within a few years, most of the public debt of the countries in crisis would be held by either official creditors or domestic banks.[4]

As a result of these features, the Portuguese crisis combined a deep recession and a debt crisis, as in other crisis countries (Fagan and Gaspar 2007; Bento 2010; Reis 2013; Alexandre and others 2014). What, then, was special about the Portuguese crisis? To start, Portugal's recession had not begun with a crash in 2010, but rather with a slump that had been going on for 10 years before that. In the 2000–09 period, real GDP per capita had grown by only 2.9 percent and the unemployment rate had risen from 4.9 percent to 11.3 percent. The extent of the economic calamity in Greece during the crisis has been often emphasized: Greek real GDP grew cumulatively by only 1.4 percent between 2000 and 2012. But Portugal grew by the same 1.4 percent during the same period, because it was already slumping in the first 10 years of the century. In addition, Portugal did not have a housing price boom nor a significant expansion of its construction sector before the crisis. Instead, the large expansion in nontradables and consequent appreciation of the real exchange rate that came with the large capital inflows from the rest of Europe took place in the wholesale and retail sectors and in community services (education, health care, and social work).

The debt crisis in Portugal also had two distinct features relative to the other crisis countries. First, there is little evidence of public profligacy in Portugal before 2007. All of the increase in public spending is accounted for by increases in the payment of old-age pensions and unemployment benefits, and both of these systems actually became less generous during this period (Reis 2013). Moreover, taxes increased. Second, partly because of the contraction in income after 2000 without as large a contraction in consumption, private external debt was higher in Portugal than in the other euro-crisis countries: net international liabilities were already 104 percent of GDP by the end of 2010.

4. Fonseca, Crosignani, and Faria-e-Castro (2015) document the increase in banks' holdings of Portuguese debt.

Combining these features that were common to the euro crisis with Portugal's specific characteristics, the challenge of the adjustment program was to deal with four problems: How could one pay for the accumulated debt, public and private? How could one control public spending, especially in pensions? How could one leave the slump and restart growth? And how could one restore competitiveness by improving the allocation of resources in the economy? The next four sections assess the program through these four lenses. I leave politics out of the discussion until the conclusion because of a final Portuguese distinction in its economic and debt crisis: there was a relative political consensus around the adjustment programs. Unlike what happened in Greece and Spain, the programs were signed onto by the three major center parties, and their share of the votes in polls fell only slightly during the program without handing new or radical parties large gains.

II. Paying Past Debts

At the start of the program, Portugal had both large public debt and large external debt. Both the private and the public sector experienced difficulty rolling over these debts, and debt overhang was holding back new investment, so adjustment required dealing with this debt.

Concerning paying the national debt, the trade balance went from −7.6 to 0.5 percent of GDP.[5] The country had not had a trade surplus since World War II, so this was no small accomplishment. At the same time, if this had been achieved through a contraction in imports, both because of a contraction in aggregate demand during the crisis and because the price of oil fell, one might worry that this improvement was temporary. Arguing against this is the fact that the ratio of exports to GDP increased from 29.9 to 39.9 percent, as well as the fact that Portugal improved its share in most of its export markets.

Turning attention to the public debt, the budget deficit improved from −11.2 to −7.2 percent of GDP. Part of this was due to the reduction in interest payments when privately held debt was rolled over into troika debt. Nevertheless, the primary surplus also improved markedly, rising from −8.2 to −2.3 percent of GDP. Much ink has been spilled on the virtues and pitfalls of austerity in a debt crisis. One interesting feature of the Portuguese situation (and the euro crisis) is represented in table 1. Aside from the primary surplus since 2010, the table also shows its projected

5. All comparisons are between 2010 and 2014, using annual data, unless stated otherwise.

Table 1. Public Primary Deficits: Actual and IMF Forecasts

		Forecasts		
Year	*Actual*	*June 2011*	*October 2012*	*January 2013*
2005	−3.6			
2006	−1.6			
2007	−0.1			
2008	−0.7			
2010	−8.2			
2011	−3.1	−1.7		
2012	−0.8	0.3		
2013	0.0	2.1	0.2	−0.2
2014	2.3	2.8	2.4	2.1

Sources: IMF reports on Portugal for the years given above.

path according to different waves of IMF programs, as well as the fall in the deficit between 2005 and 2008, when Portugal was in violation of the Maastricht limits and had to bring its deficit in line. The pace of austerity was milder than what had been planned, with constantly relaxed targets, and it was similar in 2011–14 to what it had been in 2005–08. It is hard to make the case that there was unexpected austerity from the start of 2012 onwards, or to see a dramatic reform in Portuguese public finances. Another jarring comparison is that made with the United States. Between 2010 and 2014, the U.S. federal surplus improved by 5.9 percent, in spite of little talk of excessive austerity and no troika impositions; Portugal's surplus improved by only 4.0 percent.

Lowering the debt can also be done by selling assets or by restructuring liabilities. A sign of the first activity is evident in Portugal's gross external debt, which grew by only €0.5 billion, while its net international investment position worsened by only €8.3 billion. Large companies, both public and privately owned, were sold to foreigners, including the major electrical utility, the larger telecommunications company, the airline, and large banks.

As for the second, in spite of the cut in the public deficit, the stock of public debt went from 96 percent to 130 percent of GDP. This number may be misleading because it refers to the face value of the debt. However, in 2012, the troika restructured the debt of Portugal (together with that of Greece and Ireland), extending maturities and lowering interest payments, thereby reducing its market value in spite of making no cuts to the face value. Since a large share of the debt is owed to the troika institutions and is not traded, there is no market value to assess. Following Daniel Dias, Christine Richmond, and Mark Wright (2014) and Julian Schumacher and

Beatrice Weder di Mauro (2015), I calculate the present value of the payments that the Portuguese government has committed to make to all of the holders of its debt, both private and public. If, following these authors, one uses a subjective interest rate of 5 percent per year to discount the payments, then the market value of the debt is 80 percent of its face value. If instead one uses the market discount rates for the yield curve on Portuguese debt, its market value is 95 percent of its face value.

Either way, Portugal still has a high public debt outstanding and a meager primary surplus. It is difficult to see how Portugal can get public debt under control without a new reconfiguration of maturities and interest payments on the troika debt that more significantly reduces the market value of the public debt. The radicalization of European public opinion caused by the 2015 Greek crisis has made this harder to achieve.

III. Getting Public Spending under Control

A large part of the reduction in the public deficit was achieved by increasing the tax rates on personal income and sales as well as tighter enforcement. Overall government revenue increased from 40.6 to 44.5 percent of GDP. At the same time, government consumption purchases fell from 20.7 to 18.5 percent of GDP, and the cut in public investment was even sharper, from 5.3 to 2.0 percent.

Nevertheless, transfers increased from 22.1 to 23.2 percent of GDP, despite all the increase in public spending between 2000 and 2007 already being entirely accounted for by increases in old-age pension payments. As a result, while total public spending fell from €93 billion to €90 billion, spending excluding public investment actually rose from €84 billion to €86 billion. Once public investment returns to its precrisis levels, public spending will be almost unchanged, mostly because of the increase in social transfers.

It is up for debate whether public spending should keep on increasing, accompanied by even higher taxes. It is more clear that given current trends, the pension system in Portugal will accumulate ever larger deficits and, absent reform, be responsible for any future fiscal crises. During the adjustment programs, the retirement age increased to 66 years, early retirement was suspended, and survivor pensions became means-tested, but most reforms were either modest or generously grandfathered. The more meaningful impact on spending came from an across-the-board cut in pensions. But these cuts were partly reversed by the constitutional court, and

all the political parties have promised to fully reverse them in the next two years. As a result, a permanent adjustment to public spending is a task that mostly remains to be done.

IV. Making Structural Reforms: Growth and Labor Markets

Between 2000 and 2009, Portugal's real GDP per capita grew 7.3 percent less than the European Union average, excluding Germany. By the end of the adjustment program, in 2014, Portugal was growing 0.3 percent faster than the European Union, excluding Germany, and the IMF forecasts that it will continue doing so in the near future. Since 2000, Germany has been unusual in comparison with the rest of the European Union slumping in the beginning of the century and booming after 2010 when the rest of Europe was in crisis. Treating the European Union, excluding Germany, as the appropriate comparison, growth seems to have resumed in Portugal, starting the process of catching up to the rest of Europe.

In this comparison, it is important to note that the economic outlook is still dismal. Growth forecasts from the IMF for the next 3 years are a modest average of 1.5 percent per year, reflecting the economic stagnation of the European Union. But from the perspective of the adjustment program, it is a good sign that Portugal has resumed convergence with the rest of Europe after diverging since the start of the century.

Moreover, unlike in other European labor markets, Portuguese unemployment has fallen quite rapidly so far. The seasonally adjusted unemployment rate at the end of 2014 was 13.6 percent, still above the value at the end of 2010 (12.2 percent), but it fell to 12.2 percent by September 2015 after falling almost monotonically from its peak of 17.5 percent in January 2013.

Why this quick adjustment? One distinguishing feature of the Portuguese labor market is its dual nature (Centeno and Novo 2012). On the one hand, many workers benefit from protected contracts that make layoffs expensive, contribute to low job creation and destruction, and encourage low labor productivity. On the other hand, as many as half of all workers are on term contracts and switch jobs often. These include the large majority of jobs created in this century and are mostly held by people younger than 40. This duality is a development problem, since it lowers average productivity and makes reforms difficult. A large share of the population is unproductive and almost impossible to fire. Yet, at the margin, it implies that the Portuguese labor market is actually somewhat flexible. Because the marginal

worker is in a term contract, job creation and destruction are easy, and the unemployment rate can adjust quickly to major macroeconomic shocks.[6]

This duality also suggests that evaluating the adjustment in the labor market requires looking at the composition of employment. This will let us see whether churn in the labor market during the adjustment programs affected the average worker and average productivity in the economy. Total employment fell from 4.867 million to 4.492 million workers, although looking across sectors one finds that agriculture and construction, combined, account for two-thirds of this reduction. Looking at education levels, employment of workers with a primary school education or less declined by 824,000, so employment among those with secondary schooling or higher education actually *increased* during these four years of crisis, by 192,000 and 293,000, respectively.

Another sign of this compositional adjustment comes from the adjustment of labor compensation. Wages fell the most (by 8 percent) for those with higher education, while they fell only slightly (by 1 percent) for those with primary education or less. At the relevant margin of adjustment, wages adjusted flexibly, and employment rebounded. In aggregate, real unit labor costs fell by 6.6 percent during these four years, mostly due to a fall (5.3 percent) in real compensation.

These numbers suggest structural changes in the Portuguese economy and, perhaps, a reversal of the misallocation of resources that had plagued it for the past 15 years.

V. Competitiveness and the Allocation of Capital

In the World Economic Forum's global competitiveness index, Portugal improved in its ranking from 46th to 36th between the 2010–11 report and the 2014–15 report (Schwab 2010, 2014). This was the result of many legal reforms that were part of the extensive adjustment programs. The IMF (2015b) documents 494 different structural reform actions, about half of them in the public sector and half in the deregulation of product, labor, and financial markets. Whether any of them leads to higher economic growth is an open question.

6. Some of the decline in unemployment was certainly also due to emigration: the population fell by 172,000, or 1.6 percent. This decline in population has been steady since 2010, though, while unemployment rises and falls. A third driver of the fall in unemployment is decline in participation by the discouraged long-term unemployed, but careful statistical work to quantify how large this population was remains to be done.

Competitiveness is often measured using a real exchange rate. Yet the movements in the Portuguese effective real exchange rate were mostly due to changes in the value of the euro against other currencies. Most of the capital flows happened within European Union borders, toward nontradable sectors in the periphery, and to less productive and more protected industries (Reis 2013). A more appropriate diagnosis of competitiveness than the real exchange rate is the relative price of nontradables. Between 2010 and 2014, that price fell by only 2.4 percent, signaling little improvement.

At the same time, as noted already, exports and the tradable sector expanded considerably. The current account surplus went from −10.1 to 0.6 percent of GDP, suggesting a marked improvement in competitiveness. Much as in the years before the crisis, there was a significant reallocation across tradables and nontradables during the adjustment in spite of small changes in relative prices.

To find signs of an improved allocation of resources, it would be desirable to have estimates of productivity and markups. Neither type of estimate is available with current data. However, a much more imperfect measure of labor productivity—output per hour—is available: it increased by 2.8 percent in the overall economy. More interestingly, Reis (2013) emphasizes that two sectors—retail and wholesale trade and real estate activities—had a large increase in markups and stagnant productivity in 2000–07 and yet absorbed large amounts of the capital inflow. These two sectors had among the largest increases in output per hour between 2010 and 2014—11.1 percent in wholesale and retail trade and 10.8 percent in real estate—even as they shrank in their relative size. This evolution is consistent with misallocation and inefficiency before the crisis and with an improvement during the adjustment program.

As is typical in Europe, the financial system is dominated by banks, which are crucial in allocating capital across sectors. Since 2010, the shareholders of most banks in Portugal lost almost all their investment after several waves of recapitalization, with one of the four major banks going through resolution. Moreover, the banks were subject to the ECB's asset quality reviews as well as more intense regulation. The fall in total loans is more than fully accounted for by the decline in loans to the construction sector, and the ratio of credit to deposits increased.

There are reasons to be wary of the state of banks' finances. First, nonperforming loans to nonfinancial corporations have increased almost continuously, from 4 to 14 percent. Rather than rising sharply at the start of the adjustment program as banks and regulators revalued assets, this slow

and prolonged acknowledgment of losses suggests that banks may have been rolling over bad loans. Second, and confirming this fear, corporate debt stayed almost unchanged at 153 percent of GDP. By comparison, in Spain during the same period, corporate debt fell by more than 20 percent of GDP. The IMF (2015a) partly attributed this absence of deleveraging in Portugal to the lack of legal reforms allowing for corporate bankruptcies and debt write-downs.

VI. Conclusions

It is difficult to call an adjustment process a success when the country in question has barely grown in 15 years and unemployment is 12.2 percent, yet the Portuguese economy has changed in many directions that seem promising. The misallocation of resources that plagued it seems to have started to reverse, as export sectors have grown, employment has shifted to more educated workers, the protection of local interests has declined, and output per hour has increased in the least productive sectors. The economy is growing faster than in the rest of Europe. While the definitive tests of adjustment will be whether economic growth in the next few years is able to offset the stagnation of the last 15 years, there are encouraging signs of success.

At the same time, it is easy to claim success in adjusting public finances when looking at the profile of stable and small payments that the Portuguese state has to make in the near term. However, behind the low interest rates and longer maturities, public debt is 130 percent of GDP, austerity was far from being decisive in generating large primary surpluses, and public spending will keep on rising given the lack of pension system reform. The evolution of public finances is closer to being a failure than a success, and without a quiet restructuring of the debt to the European authorities over the next few years that could lower its market value, there are reasons to be worried.

In the near term, as the recent Greek crisis illustrates, it is often politics that derails adjustment. In this regard, the troika has found a very committed and cooperative government in Portugal during the last four years. At the same time the troika insisted on changes in pensions, which have repeatedly been deemed unconstitutional by the courts, it pushed for changes to the structure of payroll taxes that were very unpopular, and it sent contradictory public messages on the need to adjust public finances. Starting from an initial position of support for reforms, the troika made

itself unpopular, often unnecessarily. Even if there is no dramatic reversal of the reforms so far, it is uncertain whether what remains to be done will ever take place.

ACKNOWLEDGMENTS I am grateful to the editors for discussions and encouragement, to Kevin O'Rourke for his discussion, and to Cynthia Balloch for research assistance.

References

Alexandre, Fernando, Pedro Bação, Pedro Lains, Manuel M. F. Martins, and others, editors. 2014. *A Economia Portuguesa na União Europeia* [The Portuguese Economy in the European Union]: *1986–2010.* Lisbon: Actual Editora.

Bento, Vitor. 2010. *O Nó Cego da Economia: Como Resolver o Principal Bloqueio do Crescimento Económico* [The Blind Knot of the Economy: How to Solve the Main Blockage of Economic Growth]. Lisbon: Bnomics.

Blanchard, Olivier. 2013. Comment to "The Portuguese Slump and Crash and the Euro Crisis." *Brookings Papers on Economic Activity,* Spring: 194–97.

Brunnermeier, Markus K., Luis Garicano, Phillip R. Lane, Marco Pagano, and others. 2011. "European Safe Bonds (ESBies)." http://www.columbia.edu/~rr2572/papers/11-ESBies.pdf

Brunnermeier, Markus K., and Ricardo Reis. 2015. "A Crash Course on the Euro Crisis." Presentation. http://scholar.princeton.edu/sites/default/files/01b%20EuroCrashCourse_slides_0.pdf

Centeno, Mário, and Alvaro A. Novo. 2012. "Segmentation." *Economic Bulletin* [Banco de Portugal] 18, no. 1: 7–27.

Dias, Daniel A., Christine Richmond, and Mark L. J. Wright. 2014. "The Stock of External Sovereign Debt: Can We Take the Data at 'Face Value'?" *Journal of International Economics* 94, no. 1: 1–17.

European Commission. 2014. "The Economic Adjustment Programme for Portugal 2011–2014." Occasional Paper no. 202. Brussels.

Fagan, Gabriel, and Vitor Gaspar. 2007. "Adjusting to the Euro." Working Paper no. 716. European Central Bank.

Fonseca, Luís, Matteo Crosignani, and Miguel Faria-e-Castro. 2015. "Central Bank Interventions, Demand for Collateral, and Sovereign Borrowing Costs." Working Paper no. 9. Banco de Portugal.

German Federal Ministry of Finance. 2014. "Portugal Exits IMF-EU Programme/Finance Minister Schäuble Welcomes Decision." http://www.bundesfinanzministerium.de/Content/EN/Standardartikel/Topics/Europe/Articles/2014-06-12-portugal-exits-imf-eu-programme.html

IMF (International Monetary Fund). 2015a. "Portugal: First Post-Program Monitoring Discussions—Staff Report; Press Release; and Statement by Executive Director." IMF Country Report no. 15/21. Washington.

———. 2015b. "Portugal: Selected Issues." IMF Country Report no. 15/127. Washington.

Jorge, Rui Peres. 2014. *Os 10 erros da Troika em Portugal: Austeridade, Sacrifícios e Empobrecimento; as Reformas que Abalaram o País* [The 10 Mistakes of the Troika in Portugal: Austerity, Sacrifices, and Impoverishment; Reforms That Shook the Country]. Lisbon: A Esfera dos Livros.

Krugman, Paul. 2015. "Europe's Many Economic Disasters." Op-ed. *New York Times,* July 3.

Obstfeld, Maurice. 2013. "Finance at Center Stage: Some Lessons of the Euro Crisis." Economic Paper no. 493. Brussels: European Commission.

Reis, Ricardo. 2013. "The Portuguese Slump and Crash and the Euro Crisis." *Brookings Papers on Economic Activity,* Spring: 143–93.

———. 2015. "Gerir a Dívida Pública" [Managing Public Debt]. In *Afirmar o Futuro: Políticas Públicas para Portugal* [Affirm the Future: Public Policies for Portugal], edited by Paulo Trigo Pereira and Viriato Soromenho Marques. Lisbon: Fundação Calouste Gulbenkian.

Schumacher, Julian, and Beatrice Weder di Mauro. 2015. "Greek Debt Sustainability and Official Crisis Lending." In present volume of *Brookings Papers on Economic Activity.*

Schwab, Klaus. 2010. "The Global Competitiveness Report 2010–2011." Geneva: World Economic Forum.

———. 2014. "The Global Competitiveness Report 2014–2015." Geneva: World Economic Forum.

Shambaugh, Jay. 2012. "The Euro's Three Crises." *Brookings Papers on Economic Activity,* Spring: 157–211.

Comments and Discussion

COMMENT BY

KEVIN H. O'ROURKE Portugal burst into the news recently, following the October 4, 2015, legislative election, and the postelection maneuverings that ensued. Up until then, however, it had been the forgotten country of the eurozone crisis. Portugal's Great Recession lacks the dramatic tragedy of its Greek counterpart, or the feel-good factor of the Irish recovery. It provides less obvious arguments, both to those who support the current troika-enforced policy mix in the eurozone periphery (if only the Greeks were more like the Irish, their economy would be in better shape) and to those who condemn it (the decline in Greek GDP since 2008 has been Great Depression–sized, while even in Ireland the post-2008 experience was traumatic). And so we have tended to ignore Portugal, which is probably a mistake.

The root cause of Portugal's ejection from international capital markets, and subsequent entry into a bailout program, was slow economic growth over many years. Between 1999 (when the euro was introduced) and 2008, Portuguese GDP per capita grew by just 9.4 percent, or 1 percent per annum. This is unlike the experience in the other crisis countries: Over the same period, per capita GDP grew by 31 percent in Ireland, 33 percent in Greece, and 19 percent in Spain. Portugal's postcrisis downturn was also much less exciting than elsewhere in the periphery. Greece's per capita GDP in 2014 was just 75 percent of its 2008 level; the comparable figures for Spain and Ireland are 92 percent and 101 percent respectively. By contrast, the figure is 94 percent in Portugal, neither as bad as in the basket cases, nor as "good" as in Ireland. In short, Portugal is a bit boring.[1]

1. Here and elsewhere, per capita GDP data are taken from the AMECO database (http://ec.europa.eu/economy_finance/db_indicators/ameco/index_en.htm), series RVGDP.

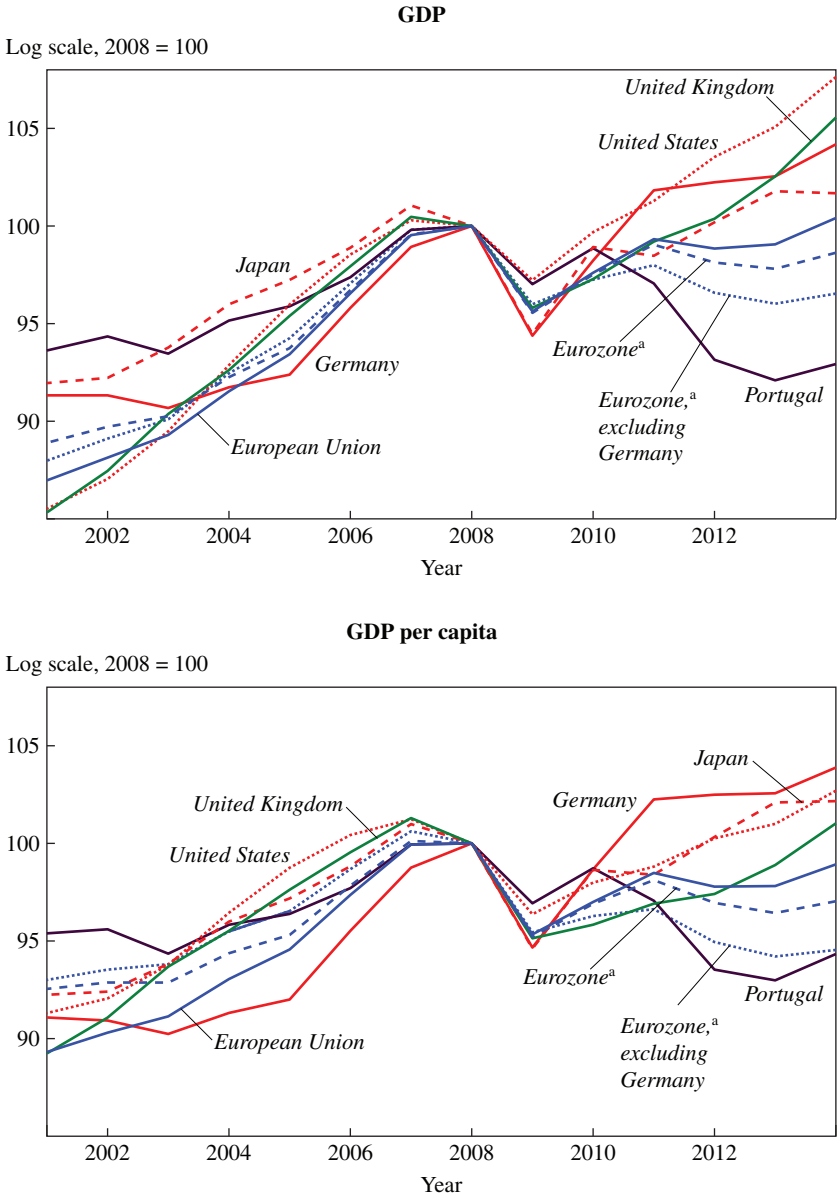
But this is precisely why we should pay more attention to the country. Greece and Ireland are both sui generis: on the one side, a corrupt and fiscally profligate government; on the other, one of the biggest housing bubbles and busts in history. In contrast, Portugal's ploddingly slow growth is much more akin to what other eurozone economies have experienced, albeit without falling into bailout programs. In France, per capita GDP rose by 11 percent between 1999 and 2008, while in Italy it rose by just 7 percent (or by three quarters of a percentage point per annum). The Portuguese experience falls squarely within this range, making it a fairly canonical example of an underperforming eurozone economy. Its misfortune, perhaps, was that it fell into difficulties while Jean-Claude Trichet rather than Mario Draghi was at the helm of the European Central Bank in Frankfurt (or, perhaps, that it was not viewed as sufficiently important to trigger a "whatever it takes" declaration, unlike Spain and Italy).[2]

Ricardo Reis's paper provides an admirably nuanced account of Portugal's performance since the troika took over in 2011. He argues that while the fiscal problems that plagued the country in the run-up to the crisis, due essentially to a combination of slow growth and rising entitlement spending, persist, valuable progress has been made in rebalancing the real economy. If the root cause of the problem was slow growth, however, restoring growth would seem to be key to resolving it, and Reis acknowledges that "the economic outlook is still dismal." He draws comfort, however, from the fact that Portuguese growth is now faster than the European Union (EU) average, once Germany has been removed from the equation, suggesting that convergence has finally resumed.

Reis's paper is looking for a eurozone success story. Unfortunately, if you are looking for success in the eurozone you are looking in the wrong place. My figure 1 plots per capita and aggregate GDP from 2001 to 2014 for the United States, United Kingdom, European Union, eurozone, Germany, Japan, and Portugal. For the sake of consistency over time, I look at just the original 12 eurozone members, including Greece, which joined in 2001. I also plot the data for the eurozone thus defined, minus Germany, following Athanasios Orphanides (2015). The figure makes it clear how utterly woeful Europe's performance has been since the crisis. In per capita terms, neither the EU nor the eurozone had recovered to the precrisis peak by 2014, in stark contrast with the United States (whose recovery is usually viewed as disappointing by Americans) and Japan. In aggregate terms, GDP in the EU only just recovered to its precrisis level in 2014, while

2. Very high levels of private indebtedness may also have played a role.

Figure 1. GDP and GDP per Capita, 2001–14



Source: European Commission, AMECO database, series RVGDP, OVGD, and NPTD (http://ec.europa.eu/economy_finance/db_indicators/ameco/index_en.htm).

a. Refers to the 12 original member states of the eurozone: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, and Spain.

eurozone GDP remained 1.4 percent below it (and U.S. GDP was 7.6 percent above).

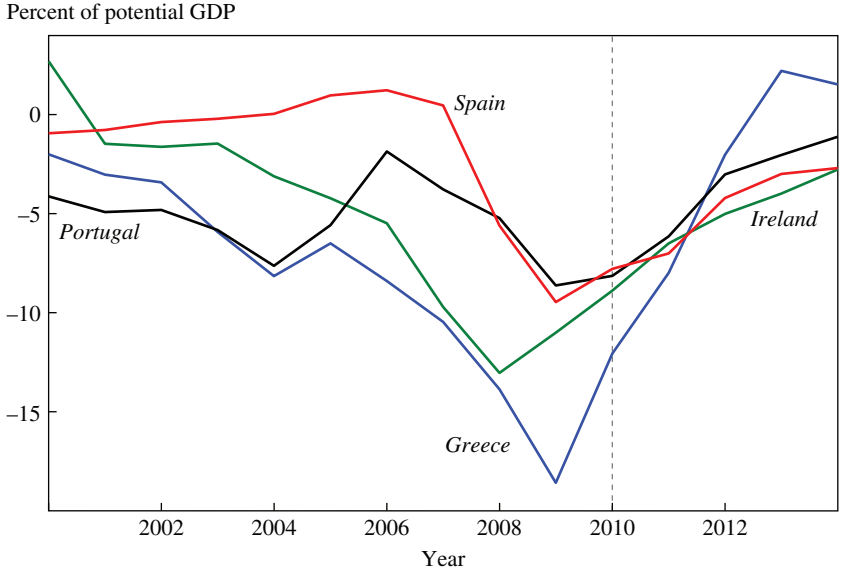
My figure 1 shows that it does indeed make sense to distinguish between the performances of Germany and the rest of the eurozone. In per capita terms, Germany's recovery has surpassed that of the United Kingdom, and even the United States; per capita GDP was almost 4 percent higher in 2014 than in 2008. In the 12 original eurozone member states, 2014 per capita GDP was almost 3 percent lower than in 2008; excluding Germany, it was 5.5 percent lower. Strikingly, Portugal's per capita GDP is now about the same, relative to the precrisis peak, as in the rest of the non-German eurozone. This is due to the more rapid growth in 2014 alluded to in Reis's paper, and is clearly visible in the graph. Whether this improvement in Portugal's relative performance is more a reflection of Portuguese success or eurozone failure is another matter. For my part, I look at my figure 1 and see a European performance that is truly dismal.

Also worth mentioning is the fact that Portugal's relative performance is a lot less impressive when expressed in terms of aggregate rather than per capita GDP. This discrepancy is in part due to emigration, which as Reis notes has led to Portugal's population falling since 2010. Conversely, emigration has presumably played a role in keeping a lid on Portuguese unemployment. In 2012 the International Monetary Fund (IMF) calculated that Ireland's unemployment rate would have exceeded 20 percent had it not been for the emigration safety valve, and it would be interesting to see a similar calculation performed for the Portuguese case (IMF 2012, p. 5).

In summary, it is hard to feel much optimism regarding Portugal's recent growth experience, and its growth prospects are not that great either. Kieran McQuinn and Karl Wheland (2015) estimate that the eurozone will grow at 0.6 percent per annum between 2014 and 2023, and at 0.25 percent or less for the subsequent 20 years. Converging on such a laggard is nothing to be happy about: They estimate that Portugal will grow at just 0.81 percent per annum between 2014 and 2023, at 0.08 percent per annum between 2024 and 2033, and at −0.32 percent per annum between 2034 and 2043. If they are right, then Portugal's fundamental problem, and the root cause of the 2011 bailout, is far from being solved.

If Portugal does not offer clear evidence of a troika success, what about Ireland? While no one should believe Ireland's GDP (or even GNP) data (FitzGerald 2013), there has been impressive growth in the country since 2013, as evidenced in rising employment numbers. But this should not be taken as evidence that the troika's intervention in Ireland "worked," or

Figure 2. General Government Structural Balances, 2000–2014[a]



Source: International Monetary Fund, World Economic Outlook Database (http://www.imf.org/external/pubs/ft/weo/2015/01/weodata/index.aspx).

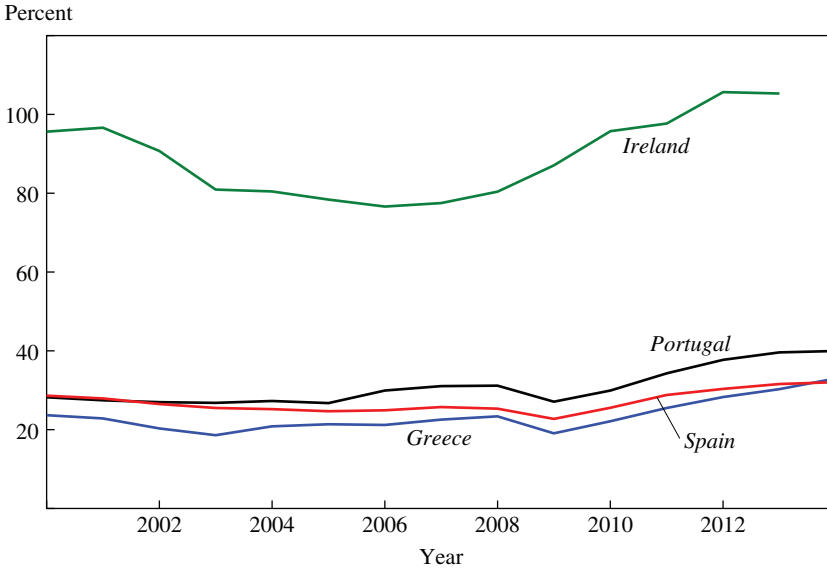
a. The structural budget position is defined as the actual budget deficit (or surplus) less the effects of cyclical deviations of output from potential output.

that its medicine, if taken enthusiastically, will produce similar recoveries elsewhere in the eurozone.

Ireland's fiscal consolidation, while impressive, preceded the troika's arrival in the autumn of 2010 (my figure 2), so it is hard to either credit or blame the troika for Irish austerity. More importantly, perhaps, it is not surprising that there was political buy-in for an austerity program that was decided domestically, rather than being imposed from abroad. Another factor that must have helped in getting austerity through the political system was that Irish incomes were so high to begin with; on the eve of the crisis, per capita GDP was more than twice as high in Ireland as in Portugal, and almost twice as high as in Greece. Even subtracting 20 percent from the Irish numbers, to account for the distinction between GDP and GNP, does not eliminate the reality that Ireland is a very rich country. This surely made it easier to cut people's incomes.

My figure 2 shows that Ireland suffered considerably less austerity than Greece (but more than Portugal). Ireland is also a *much* more open economy than any of the other crisis countries (my figure 3), and is famously

Figure 3. Exports as a Percent of GDP, 2000–14



Source: World Bank, *World Development Indicators.*

dependent on inward foreign direct investment. This surely implies lower multipliers in Ireland than in either Greece or Portugal. Ireland also benefited from the fact that its major trading partners, the United Kingdom and United States, were outside the eurozone. This should have been a good reason for Ireland to stay outside the euro, but ironically enough it turned out to be a saving grace when the crisis hit. As my figure 1 showed, those countries' postcrisis performances, while disappointing, were streets ahead of the eurozone's. Ireland was therefore exporting into growing markets, and also benefited more than other crisis countries from the euro's depreciation in 2014 and 2015. Between January 2014 and July 2015, Ireland experienced an 8.7 percent nominal (trade-weighted) depreciation. The equivalent figures in Greece, Spain, and Portugal were 4.3 percent, 4.5 percent, and just 2.9 percent. Nominal depreciations translated into real depreciations: in Ireland's case, 9.2 percent, as opposed to an impressive 7.8 percent in Greece, 6 percent in Spain, and 2.7 percent in Portugal.[3] Not surprisingly, 2014 was when Irish growth really picked

3. Source: Bank for International Settlements' effective exchange rate indexes (http://www.bis.org/statistics/eer.htm).

up. Interestingly enough, in percentage terms Portugal's exports have been growing at rates roughly comparable to Ireland's, the difference being that a given percentage increase in exports will have a bigger impact if exports are around 100 percent of GDP than if they are less than 40 percent. But Portuguese export growth remains impressive, not least because the country benefited from smaller nominal and real depreciations than other crisis countries. The apparently high price elasticities that this seems to imply, and that Reis alludes to in his paper, are interesting and seem worthy of further study.

Was the Irish success story due to the growth-promoting effects of troika-inspired structural reforms? It seems unlikely. True, Ireland's record when it came to implementing structural reforms was better than either Greece's or Portugal's; "close to perfect" is how Alessio Terzi (2015, p. 5) describes it. Conversely, the same author shows that the *number* of reforms that Ireland implemented was much lower than in Portugal or Greece, and that Ireland's reform efforts overwhelmingly involved the country's financial sector, rather than politically difficult areas such as pensions or labor market reforms (Terzi 2015, p. 4). Given that Ireland entered the crisis with a relatively flexible economy, there were fewer growth-promoting structural reforms that could be implemented there, making it difficult to attribute Ireland's recovery to the implementation of such reforms.[4]

To summarize: Austerity in Ireland was home-grown; it was less extreme than in Greece; and its impact was lower because of Ireland's high incomes, its extreme openness, and the emigration safety valve. Structural reforms probably did not boost growth much if at all, but Ireland got lucky, benefiting from growing trade partners and a depreciating currency. Its economy has been growing rapidly for two years. Despite all these advantages, the Irish political landscape has been turned on its head. In the 2011 general election, the three main parties (Fianna Fáil, Fine Gael, and Labour) obtained 73 percent of the vote, down from 79 percent in 2007. They are now standing at around 57 percent in the polls, following

4. McQuinn and Whelan (2015) try to derive an upper bound estimate of the impact of pension, labor market, and regulatory reforms on European growth, assuming inter alia that regulatory reforms would raise total factor productivity to U.K. levels by 2043. Even on this heroic assumption, structural reforms would only add 0.85 percentage point to eurozone growth between 2014 and 2043, implying growth of 1.2 percent per annum. Interestingly, they estimate that the same reforms would add as much as 1.58 percent per annum to Portuguese growth between 2014 and 2043, yielding an average growth rate of 1.81 percent per annum.

a big swing toward Sinn Féin and an assortment of mainly left-wing protest candidates.[5] European policymakers might want to think twice before imposing such a policy mix on larger, more closed economies without Ireland's enviable history of political stability.

REFERENCES FOR THE O'ROURKE COMMENT

FitzGerald, John. 2013. "The Effect of Redomiciled Plcs on GNP and the Irish Balance of Payments." Research Note no. 2013/1/2. In *Quarterly Economic Commentary,* Spring, by David Duffy and Kevin Timoney. Dublin: Economic and Social Research Institute.

IMF (International Monetary Fund). 2012. "Ireland: Eighth Review under the Extended Arrangement; Staff Report; Staff Supplements; and Press Release on the Executive Board Discussion." Country Report no. 12/336. Washington.

McQuinn, Kieran, and Karl Whelan. 2015. "Europe's Long-Term Growth Prospects: With and Without Structural Reforms." Working Paper no. 15/08. Dublin: University College Dublin, Centre for Economic Research.

Orphanides, Athanasios. 2015. "The Euro Area Crisis Five Years after the Original Sin." Working Paper no. 5147-15. Cambridge, Mass.: Massachusetts Institute of Technology, Sloan School of Management.

Terzi, Alessio. 2015. "Reform Momentum and Its Impact on Greek Growth." Policy Contribution no. 2015/12. Brussels: Bruegel.

GENERAL DISCUSSION Beatrice Weder di Mauro opened the discussion by noting that Ricardo Reis seemed more optimistic about Portugal's success in his conference presentation of the paper than in the paper itself. Portugal's recovery story is certainly different from that of Greece or Ireland, but whether one should call it a success depends on the criteria, most of all on what one is comparing it to. The paper treats growth in the euro area, minus Germany, as the counterfactual, but that did not strike Weder di Mauro as the most meaningful comparison. It would make more sense to compare Portugal's experience with that of other countries undergoing IMF-type programs, making use of the kind of evaluations the IMF does periodically when adjustment programs are put into place.

Steve Davis had three questions. First, how bad is Portugal's pension reform problem—that is, what is the projected path for pensions as a percentage of GDP without reform? Second, what has happened in the

5. Recovery does seem to be helping the centrist parties; the figure is up from just 46 percent in December 2014. See http://www.oireachtas.ie/parliament and http://www.redcresearch.ie/election-2016.

retail and wholesale sectors to facilitate the country's recent productivity improvements? And third, can the labor-productivity measures for Portugal be trusted? Davis's concern on this last point was that the output and labor input measures in Portugal, particularly in retail, are differently sensitive to unreported activity, and unreported labor inputs are easier to disguise than sales numbers. If hiring has been occurring off the books, and especially if such activity increases or decreases, that will distort the evolution over time of labor-productivity measures.

Robert Gordon wondered why Reis's paper only showed ratios to GDP when measuring whether transfers went up or down, pension costs went up or down, and the same for expenditures and exports. After all, GDP in Portugal has fallen. Would it not be more useful to measure ratios to potential output or to previous peak output, so that one could see whether these things had gone up or down in absolute terms instead of relative to a shrinking total?

Yannis Ioannides wondered about the profile of Portugal's payment obligations. In the paper, Reis predicted that without a quiet restructuring of its debts to the European authorities over the next few years, in a way that would lower their market value without affecting their face value, there would be reasons to worry. Ioannides agreed with Reis that accomplishing this in the Greek crisis was much harder, yet he wondered whether the profile of Portugal's payment obligations was so significantly different that it would be easier there. The payments it owed to the European Central Bank and the International Monetary Fund in 2015, for example, did seem substantial.

Jay Shambaugh was curious to know whether Reis regarded Portugal's experience as a macroeconomic *policy* success. He agreed that the shift away from nontraded goods and the fall in tradable unit labor costs have been impressive, but he wondered whether the transition needed to be as painful as it was. What if the European Central Bank had been more supportive as early as 2011, or if there had been supportive fiscal policy across the eurozone as more of a single macroeconomic unit? Then rather than arguing over whether Portugal's austerity itself was bad one might have wound up with a better recovered euro area, excluding Germany, instead of an area down 6 percent GDP per capita.

Reis responded by addressing five issues that had been directly or indirectly raised. First, concerning long-run growth, he believed the prospects for the entire eurozone look dismal and the same must be said for Portugal. He considered the country to have had a macroeconomic success only insofar as it weathered an adjustment well. It is even doing a bit better than

Italy, one of its comparator countries, which had no adjustment program. His point in the paper was not to say Portugal is now doing well in absolute terms and can be called a great success story, nor even, in response to Shambaugh's comment, that there have been no policy mistakes.

A second issue concerned the labor market. As in Italy and Spain, Reis said, in Portugal a large fraction of the population is accustomed to extremely protected labor markets that keep many workers from being fired under almost any circumstance. In response, the economy has ended up with a growing number of temporary contract workers. This problem is greatest in Portugal, where it was already evident almost 20 years ago. Of all new jobs created since 2000, 80 percent have been such temporary positions. As a result, somewhere between a third to half the labor force—depending on estimates—is now employed through these flexible contracts. The resulting dual labor markets are not a good thing. In Portugal now, 50 percent of the labor force has very low productivity, which is partly why the total factor productivity levels in growth are so low. At the same time, it has the advantage that when a shock occurs and employment must adjust up or down it allows employers a lot of discretion.

A third issue concerned the puzzle of international exchange rates and their impact on balance of trade. Usually, while exchange rates move a lot, credit accounts do not. Currency unions like the eurozone's create a special situation, a kind of hyperconnected exchange rate. Starting around 2000, some of the current accounts, including Ireland's and Portugal's, went through massive swings with very large deficits, even with moderate CPI-based rate changes, and now we see credit accounts fall tremendously. So trade balances go up and down with fairly moderate real exchange rate changes, and this is true not only for Portugal.

In response to Davis's concern about the reliability of labor input measures, a fourth issue, Reis described Portugal's new tax-compliance measures. To increase its revenues, the country has adopted IMF-recommended measures, and consequently tax compliance has improved quite a bit, especially in retail and wholesale trade. For instance, shops not only need to receive invoices but their cash registers are electronically connected to the revenue service. The sales staff are even incentivized to use the system through government lottery prizes, including expensive cars. These approaches have been extremely effective, an example of applied behavioral economics. The other benefit is that reported wholesale and retail sales have risen massively. There had been tax evasion in the past, but now, with better reporting of output, the numbers show higher output per hour as well.

Finally, Reis agreed with the point raised about immigration, that it has an important effect on employment and recovery. In addition to people leaving Portugal for work, the country has stopped receiving Eastern Europeans who left their home countries looking for jobs. Reis suspected that migration cannot explain all the upward movement of the employment rate, however. It is hard to measure this accurately, though, until the census has been taken.