

CFA INSTITUTE RESEARCH FOUNDATION / **MONOGRAPH**

THE PRODUCTIVITY PUZZLE

Restoring Economic Dynamism

**DAVID E. ADLER AND LAURENCE B. SIEGEL,
EDITORS**



**CFA Institute
Research
Foundation**

THE PRODUCTIVITY PUZZLE

Restoring Economic Dynamism

David E. Adler and Laurence B. Siegel, editors



CFA Institute
Research
Foundation

Statement of Purpose

The CFA Institute Research Foundation is a not-for-profit organization established to promote the development and dissemination of relevant research for investment practitioners worldwide.

Neither CFA Institute Research Foundation, CFA Institute, nor the publication's editorial staff is responsible for facts and opinions presented in this publication. This publication reflects the views of the author(s) and does not represent the official views of CFA Institute Research Foundation.

CFA®, Chartered Financial Analyst®, and GIPS® are just a few of the trademarks owned by CFA Institute. To view a list of CFA Institute trademarks and the Guide for the Use of CFA Institute Marks, please visit our website at www.cfainstitute.org.

© 2019 CFA Institute Research Foundation. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the copyright holder.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional service. If legal advice or other expert assistance is required, the services of a competent professional should be sought.

Cover photo credit: sivarock / E+ / Getty Images

ISBN 978-1-944960-83-4

Biographies

David E. Adler is the author of *Snap Judgment* (FT Press) about behavioral economics and a producer of the related NOVA documentary *Mind Over Money*. He also wrote *The New Economics of Liquidity and Financial Frictions* and *The New Field of Liquidity and Financial Frictions* (CFA Institute Research Foundation). He is also producer of the PBS documentary *America's Crisis in Healthcare and Retirement* and co-editor of the anthology *Understanding American Economic Decline* (Cambridge University Press), which looks at the prospects for US growth through the precedent of Britain's comparative economic decline. He is a frequent contributor to *City Journal* and *American Affairs* and is a senior advisor to XA Investments in Chicago.

Laurence B. Siegel is the Gary P. Brinson Director of Research at the CFA Institute Research Foundation and an independent consultant, writer, and speaker. In July 2009, he retired from the Ford Foundation, where he served as director of research in the Investment Division since 1994. From 1979 to 1994, he was a managing director of Ibbotson Associates. Mr. Siegel's first trade book, *Fewer, Richer, Greener: Prospects for Humanity in an Age of Abundance* (Wiley), will be released in late fall 2019. In addition, he has written, co-authored, or edited six CFA Institute Research Foundation books. He serves on the board of directors of Q Group (Institute for Quantitative Research in Finance) and the American Business History Center. He received his BA in urban geography from the University of Chicago in 1975 and his MBA in finance from Chicago Booth in 1977.

Contents

Preface.....	vii
David E. Adler	
Introduction	xvii
Laurence B. Siegel	
Opening Statement	1
Right Values and Wrong Ones for an Economy of Dynamism.....	2
Edmund Phelps	
History and Theory	6
A Blind Spot in Today's Macroeconomics?	7
Claudio Borio	
World War II, the US Recovery from the Great Depression of the 1930s, and Postwar Productivity and Prosperity	27
Richard Sylla	
The Age of Experts: A Review of Marc Levinson's <i>An Extraordinary Time</i>	43
Stephen Sexauer and Laurence Siegel	
Measurement and Mismeasurement	49
Underestimating the Real Growth of GDP, Personal Income, and Productivity.....	50
Martin Feldstein	
Challenges to Mismeasurement Explanations for the US Productivity Slowdown.....	74
Chad Syverson	
Difficulties and Diagnoses.....	101
Is America Still a Startup Nation?.....	102
Dane Stangler	
Understanding the Decline of US Manufacturing Employment	112
Susan N. Houseman	
Two Economies: The Decline of the American Worker and the Rise of Superstars.....	136
Anton Korinek	

Finding Capital in the Stock Market	142
David Blitzer	
Why Is Business Investment So Low in the United States?	153
Thomas Philippon	
Solutions and Proposals.....	165
Beyond Blue, Part 3: The Power of Infostructure.....	166
Walter Russell Mead	
Solving the Productivity Puzzle: The Role of Demand and the Promise of Digitization.....	173
Jaana Remes	
Building a Robust Apprenticeship System in the United States: Why and How?	197
Robert I. Lerman	
Reviving America's Forgotten Innovation System: Fostering US Growth through Incremental Product and Process Innovation....	220
Dan Breznitz and Peter Cowhey	
Supply Chains and Equitable Growth.....	246
Susan Helper	
Last Word	256
The Economic Sky Is Not Falling	257
Deirdre McCloskey	

Preface

David E. Adler

“Half the copybook wisdom of our statesmen is based on assumptions which were at one time true, or partly true, but are now less and less true day by day. We have to invent new wisdom for a new age.”

—John Maynard Keynes
Essays in Persuasion

Keynesian wisdom, and that of Monetarists as well, was necessary for understanding and addressing the economic challenges of their respective eras. But today most Western countries face a new set of problems: stagnant productivity growth and diminished dynamism. These changes are apparent in the United States, where labor productivity in many manufacturing sectors is actually declining.¹ But the productivity problem is not limited to the United States: According to McKinsey data presented in this volume, “Labor productivity-growth rates remain near historic lows in many other advanced economies.”²

The Productivity Puzzle: Restoring Economic Dynamism is a book of essays about this mysterious stagnation of productivity. It offers prescriptions for restoring productivity growth. The book’s central innovation is that it is interdisciplinary. Traditional macroeconomics has trouble fully explaining the productivity puzzle. This is because the sources of productivity growth often lie in a country’s specific economic practices and institutions. These institutional differences are better captured by political economists than by macroeconomists. But productivity also has macro components, including interest rate regimes and trade, which are typically missing from political or institutional analyses. That is, neither the macroeconomic nor institutional approach is adequate on

¹“Productivity and Costs by Industry: Manufacturing and Mining Industries—2017” (news release, US Department of Labor, Bureau of Labor Statistics, 19 April 2018), <https://www.bls.gov/news.release/prin.nr0.htm>.

²See Jaana Remes’s essay. She finds that “Sweden and the United States experienced a strong productivity boom in the mid-1990s and early 2000s followed by the largest productivity-growth decline, and much of that decline predated the financial crisis. France and Germany started from more moderate levels and experienced less of a productivity-growth decline, with most of the decline occurring after the crisis. Productivity growth was close to zero in Italy and Spain for some time well before the crisis, so severe labor shedding after the crisis actually accelerated productivity growth.”

its own. Together, both create a more complete picture and suggest new remedies that are overlooked by the conventional thinking in either field.

The stagnation in American productivity growth predates the global financial crisis of 2007–2009 with the implication that the United States risks becoming the new “sick man” of the developed world, analogous to the rotating position of the Ottoman Empire, Britain, France, Germany, and Italy as the sick man of Europe. Though many of the essays in the anthology are US-centric, simply because that is where the most easily accessible data are, the same interdisciplinary framework can extend to other countries. These countries may be deploying economic models that are now more successful than America’s in terms of manufacturing.

The essays in this anthology explore the practical dynamics of the contemporary US economy and other Western nations. This includes the rise of tech monopolist “superstar firms,” the decline in corporate investment, and the continuing impact of low interest rates. The contributors also delve into the statistical controversies surrounding productivity measurement.

This new analysis suggests new solutions, another focus of the book. In contrast, Robert Gordon’s important work of economic history, *The Rise and Fall of American Growth*,³ which comes in at 750 pages, contains only an 11-page policy “postscript.” The book’s concluding sentence, after the hundreds of pages of “declinism” that have gone before, is a special policy call-out for “pre-school education,” a disappointing coda. The productivity-enhancing strategies discussed in this anthology instead point to the need for new innovation *systems* in the United States focused on manufacturing, new labor *institutions* such as apprenticeships, new ways to build more equitable and robust supply chains, and a call for increased digitization.

Solving the productivity puzzle, and implementing productivity-enhancing policies, is not the only challenge facing the United States and the West, but it is a big one. It is one that requires new wisdom for a new age. As Paul Krugman has put it, “Productivity isn’t everything, but in the long run it is almost everything” (p. 11).⁴

A Road Map through This Book

The lead essay, by Edmund Phelps, winner of the 2006 Nobel Memorial Prize in Economic Sciences, locates the decline of Western economic dynamism in cultural values. Specifically, he argues that the loss of, and even revolt

³Robert Gordon, *The Rise and Fall of American Growth: The US Standard of Living since the Civil War* (Princeton, NJ: Princeton University Press, 2016).

⁴Paul Krugman, *The Age of Diminished Expectations: U.S. Economic Policy in the 1990s* (Cambridge, MA: MIT Press, 1990).

against, the modernist spirit has led to the sharp decline in innovation that he perceives as having started in the West in the crisis of the 1970s. Excessive regulation and corporate short-termism, both common explanations for low growth, are for Phelps merely further evidence of a culture and society that has lost its way. This essay also encapsulates a central theme in this book: the need to go beyond conventional macroeconomic models to fully understand what drives dynamism.

History and Theory

The opening chapter in this section, by Claudio Borio, head of the Monetary and Economic Department of the Bank for International Settlements, or BIS (a sort of central bank for central banks), offers a new macroeconomic explanation for flat productivity growth. Borio explains how credit booms, often the result of low interest rates, can cause “misallocations,” that is, excessive allocations to low-productivity sectors such as housing. The essay doesn’t precisely spell out the mechanism of what drives these misallocations, but there are several obvious possibilities. One is that a credit-driven boom increases domestic demand for consumption and housing over more productive sectors such as exportable manufactured goods.

The “blind spot in today’s macroeconomics” of Borio’s title refers to the lack of consideration by macro of these productivity differences among sectors and the deformed growth that can result from a credit boom. But the title could equally refer to the tunnel-like focus of macroeconomics itself on the real economy to the exclusion of the crucial financial sector—at least until after the crisis—and on the ongoing lack of interest by macroeconomists in the productivity puzzle. Borio’s assessment of the state of macro is of particular interest because his thinking is always ahead of conventional practice: At the 2003 Jackson Hole economic symposium he warned of growing financial instability. Economics finally recognized his insight, years too late. Now, years after the crisis, economics is lagging again, but in new ways. This time, having reversed its priorities at the wrong time, macro remains backward looking. It is more concerned with modeling crisis-related financial instability than with answering questions related to productivity, which we argue is the primary challenge in the future.

Richard Sylla focuses on the powerful role World War II played in 20th century US economic prosperity, offering a sweeping view of US economic history and the country’s varying productivity rate growth rates. Sylla questions the standard account that World War II was needed to end the Great Depression (a Keynesian argument) and argues that but for bad policies, the Depression would have ended much earlier.

He finds, however, that wartime innovations were directly related to the decades of prosperity that followed, an argument also supported by the Northwestern University economist Robert Gordon. In addition, Sylla emphasizes the effect of increased purchasing power and pent-up demand by consumers.

Perhaps most important, however, the reason World War II led to a postwar US boom is that every other major industrial country besides the United States was devastated. “The war catapulted the United States to a unique [monopolistic] position in modern economic history,” Sylla writes, and this position led to wonderfully high growth rates for a time: 2.96% annually from 1940 to 1970. Sylla finds that the average long-term US growth rate is much lower: “Since 1790 (1790–2016), economic growth in the United States has averaged 1.72% per year.”

We do not need or want another war, so we will have to find other ways of boosting productivity. Perhaps a 1.72% per year growth rate, sustained over very long periods, is all we can expect. But that might be good enough—it leads to a doubling of standards of living every 41 years. And that does not count gains from quality improvements and from the availability of new products that do not show up in the productivity data, which many economists argue are significant and contribute materially to well-being that is not measured in the GDP statistics.

Stephen Sexauer and Laurence Siegel’s article, “The Age of Experts,” is a book review originally published in *Business Economics*. They reviewed Marc Levinson’s book *An Extraordinary Time: The End of the Postwar Boom and the Return of the Ordinary Economy*, which argues that the economic boom of 1948–1973 represented “an economic golden age throughout the world.” Levinson describes the “development of macroeconomic models and their use in the postwar period by governments trying to harness the chaos of economic innovation and growth.” In other words, it was an age when experts on economic management and *dirigisme* enjoyed an exceptional degree of respect.

But Sexauer and Siegel point out that the quarter-century after World War II was only a golden age if you lived in the United States or the recently devastated lands of Europe and Japan. The rest of the world languished, and the extraordinary time for China, India, and many other developing areas is now. The extraordinary time of even poorer countries that have not yet entered the fullest flower of their development will be in the future. When studying history, you should not wear rose-colored glasses or have a home bias that blinds you to what is really going on.

Measurement and Mismeasurement

The next section, on measurement and mismeasurement, delves into the many statistical challenges in measuring productivity, particularly that of the “new economy” driven by smartphones and the almost universal use of computers in industry. Do the United States and other Western nations have a productivity problem, or just a productivity measurement problem? Our book contains essays by two adversaries in this debate, Harvard economist Martin Feldstein and Chad Syverson of the University of Chicago.

Feldstein argues that current methods aren’t well suited to measuring productivity changes that derive from the new economy. American statistical agencies have problems accounting for improvements in product quality, the introduction of new types of goods, and the rise of services in general, all of which offer great utility to consumers but the value of which is not easy to measure. His conclusion is that, at least in the United States, “the pace of productivity growth has been underestimated.”

Syverson focuses on the pronounced slowdown in US labor productivity growth between 2005 and 2015. This growth rate averaged 1.3% a year, which is less than half the growth rate of the preceding 10 years as well as of the period from 1947 to 1973. He shows that what he calls “the mismeasurement hypothesis”—the idea that productivity statistics understate the gains from new information and communications technology—cannot fully explain this slowdown. His findings rest on the fact that the slowdown is not related to information and communication technology (ICT) intensity.

Feldstein’s and Syverson’s essays, despite their differing views on productivity mismeasurement, aren’t really in conflict. Syverson is arguing that measurement problems alone cannot account for the recent sharp slowdown in labor productivity growth, but Feldstein is making a longer-term case, that there has still been a secular understatement in the rate of growth as a result of profound changes in the goods and services that are available in the market.

Difficulties and Diagnoses

In this section focusing on diagnoses, the first essay, by Dane Stangler, president and chief policy officer of Startup Genome, a firm that fosters entrepreneurship, questions the hype (and conventional wisdom) that America is currently a startup nation. He finds that “across nearly every indicator, business dynamism is waning in the United States. New business creation has fallen across the board.” Moreover, says Stangler, America’s entrepreneurial peak was actually in the 1970s. This pronounced decline in business dynamism has been found even in the high-tech sector since about 2000.

Despite these gloomy trends, Stangler is optimistic. He argues the United States may be on the cusp of an entrepreneurial boom driven by the favorable demographics of an expanded population aged 25 to 44, the peak age range for starting a business.

Susan Houseman, of the Upjohn Institute, scrutinizes the decline of manufacturing employment in the United States since 2000. Her analysis rests on deep knowledge of the way US statistical agencies calculate output. It overturns the prevailing (and paradoxical) narrative that this decline in manufacturing employment somehow stems from the strength of US manufacturing. According to this line of thinking, improvements in productivity and automation have reduced the need for manufacturing workers while overall output has remained steady. This turns out to be misleading.

Instead, as Houseman uncovers, US manufacturing output has collapsed along with employment. The reason for the misconception is the way US statistical agencies calculate growth in computers and semiconductors, the one alleged bright spot in US manufacturing. In reality, much of the growth in this sector turns out to be a statistical illusion, with the locus of production of these products shifting to Asia, Houseman writes. For Houseman, it is trade, rather than automation, that accounts for the sudden drop in US manufacturing employment starting in 2000.

Houseman's statistical work is important. It establishes a baseline for any informed discussion about the true state of US manufacturing. She concludes, "The widespread denial of domestic manufacturing's weakness and globalization's role in its employment collapse has inhibited much-needed, informed debate over trade policies."

Anton Korinek's essay tackles inequality. Korinek, of the University of Virginia and a former visiting scholar at numerous central banks, observes that there is a bifurcated dynamism in the American economy. Strong growth is limited to the top income percentiles of the population and a handful of cities, communities that he calls "superstars." How can the United States share this dynamism more broadly?

His underlying analysis is that the superstar phenomenon is driven by digital innovation, which gives rise to natural monopolies. These monopolies create distortions in the US economy: too little innovation in afflicted sectors as well as monopoly pricing.

Korinek proposes several policy solutions in response. These include an increased role for public investment (and presumably ownership) of basic research as well as ways to free up areas where tech firms currently have a monopolistic advantage, such as in information about consumers. He writes

that these interventions “would unleash a lot more entrepreneurialism and cut into those monopoly rents.”

S&P’s David Blitzer looks at the diminishing role of the stock market in raising capital, which in turn is affecting the wider economy. Blitzer details the decline in the number of publicly listed companies, the increase in typical listed company size, and the dearth of IPOs. This decline in IPOs has not been fully offset by an increase in private funding for startups, and productivity may suffer as a result.

Though stock market valuations have increased, Blitzer nonetheless presents a picture of declining dynamism in the sense that it is less common for fledgling firms to raise capital via public markets than in the past, and private markets are not a perfect substitute. He states, “Capital is leaving the market through buybacks, … being rearranged by mergers, and not being replenished by IPOs.”

A healthy public market for shares of businesses is essential for economic vitality, not just in the US but worldwide. More private and less public ownership of companies is an unwelcome trend, harking back to the days of large family-owned industries in Europe and Japan. Although some of these companies were innovative, perhaps because of their vertical integration, they tended to be monopolistic and run for the benefit of everyone but the consumer. If economic dynamism is to be renewed both in the United States and in other countries, the root causes of shrinking public ownership—whether attributable to regulation or other factors—must be addressed.

Economist Thomas Philippon of NYU identifies the drop in corporate investment as the reason productivity growth has been so anemic in the United States in the past decade. Corporate investment in the United States has declined from a historical 20 cents on each dollar earned to only 10 cents today. Philippon argues this decline in US investment stems from increased industry concentration in the United States and a decline in competition. He writes, “This reduction in competition explains close to two-thirds of the investment gap we have seen since 2000.”

An increase in regulation—except for antitrust regulation—may be underpinning this decrease in competition in the United States. Philippon states that as regulation increases, the industry ends up with fewer firms, higher profits, and less investment. American readers may be shocked to learn that European markets, according to Philippon, are now much more competitive than those in the United States, as can be seen in lower mobile phone prices and internal airfares. The reason? Regulators in Europe have been much more aggressive in promoting competition than the United States.

Solutions and Proposals

In terms of solutions, Walter Russell Mead, the foreign policy scholar, continues his analysis of the prevailing American post-World War II social model and its inability to continue to provide mass prosperity today. He suggests a solution: “infostructure.” By this he means finding a way to “harness the full power of IT to social needs” through new institutions, practices, and technology. Mead argues that investing in infostructure, even more than traditional infrastructure, will successfully spur communities to greater prosperity.

Jaana Remes, of the McKinsey Global Institute, offers a new assessment of the productivity puzzle in the United States and Western Europe. She identifies two distinct recent “waves” that have brought down labor productivity growth rates: the cresting of the ICT boom in the 1990s, and the aftereffects of the financial crisis. The good news is that there is now a third, more positive wave, according to Remes: “digitization, [which] contains the promise of significant productivity-boosting opportunities.”

However, as Remes observes, the benefits from digitization have not yet fully materialized because of barriers to adoption. Her remedy, which is similar in spirit to Walter Russell Mead’s, is a call for government as well as corporate policies that can hasten digitization and ultimately productivity growth. Remes concludes, “A dual focus on demand and digitization could unleash a powerful new trend of rising productivity growth that drives prosperity across advanced economies for years to come.”

The chapter by Robert Lerman, of the Urban Institute, discusses apprenticeships, why they are so valuable, and what can be done to build a widespread apprenticeship system in the United States. (Lerman has been a vocal and effective advocate for making apprenticeships part of the policy discussion in the United States.) Lerman argues that there is a need to move from a narrowly defined “academic only” approach to education to an apprenticeship model. Such a system benefits both firms and workers. It is widely deployed in Western industrial countries, including the United Kingdom, Australia, and Germany—but not the United States. Lerman’s forceful conclusion is that “it is past time for federal and state governments to make a genuine effort to build an extensive and high-value apprenticeship system.”

In a nuanced essay about innovation, political scientists Dan Breznitz and Peter Cowhey examine America’s innovation *system* and its deficiencies. They argue there are different types of innovation and claim the United States is lagging badly in some areas. Though the United States still excels in game-changing breakthroughs (novel-product innovation), it does not excel in making smaller manufacturing or design improvements, often initiated by line workers and middle managers—what the authors call incremental process

and product innovation. This latter type of innovation now often takes place in other countries.

While there is widespread worry that China may soon surpass the United States in scientific breakthroughs, the United States has already fallen far behind in applied innovation, in the expertise of making things. For instance, the United States lacks the ability to manufacture the Kindle domestically. Breznitz and Cowhey locate these failures in America's institutional setup, including short-termism in its financial sector, as well as lack of applied research institutions to support manufacturing, such as the Fraunhofer Institutes in Germany. They propose several remedies to revive applied innovation, particularly in manufacturing in the United States, including "network solutions" to upgrade capabilities for I&P innovation."

Case Western Reserve University economist Susan Helper, who was formerly chief economist at the US Department of Commerce and a member of the White House staff during the Obama administration, focuses on US manufacturing supply chains. They have undergone changes that have played a little known but outsized role in dampening productivity growth. Tasks formerly done in-house by vertically integrated companies are now outsourced in ways that have led to inequality and productivity stagnation in the United States. In essence, purchasing departments favor the lowest-cost supplier firms, but at the hidden cost of suppressed or eliminated innovation.

Helper lays out steps for a more collaborative approach that can lead to greater innovation and higher wages. This includes a role for federal policy: National labs could work with smaller firms or entire supply chains to share technical knowledge in manufacturing. This approach has already been successfully tried in US agriculture, where land grant universities have historically helped spread advanced knowledge about farming techniques. The vision Helper lays out is one of "high road" policies leading to good jobs for American workers and revived innovation and productivity growth in manufacturing.

Last Word

We give the last word to the philosopher and generalist Deirdre McCloskey. McCloskey ends on a needed hopeful note. As she points out, average real income around the world is rising—very rapidly, in many developing countries—and it is impossible to forecast what further technological breakthroughs might occur, and where.

The poor may always be with us but they are getting richer, according to McCloskey. Being poor means something quite different in 2018 than it did in 1818 or, in developing countries, just a generation ago. In many locales the

poor now have access to an adequate food supply, a wide variety of consumer goods, and modern medical treatments. McCloskey also carefully distinguishes absolute from relative decline.

Nonetheless, as other contributors to this book point out, future rapid productivity growth in the US and much of the West is far from guaranteed. We hope the essays in this anthology offer new thinking about the productivity puzzle. These new ideas and new policies could restore the robust productivity growth of developed countries, including both America and the rest of the West, and encourage continued inventiveness and dynamism elsewhere.

Introduction

Laurence B. Siegel

Is the world economy in a long-term slowdown? Are our best days behind us? Is productivity growth, widely agreed on as the principal driver of broader economic growth, headed for a long period of subpar performance? If so, what can we do to remedy the situation and restore economic dynamism?

Investors, business leaders, policymakers, and academic economists each have their perspectives on these questions—and all of them need to know the answers, as best we can determine them. In a landmark conference organized by David E. Adler and me, and financially supported by the CFA Institute Research Foundation, Standard & Poor's, and McKinsey & Company, thinkers from a wide variety of backgrounds and organizations addressed these issues. The conference took place at the Museum of American Finance in New York on 28 November 2017.

Thinking Differently

The conference was organized along interdisciplinary lines. People in the same field are often subject to a herd instinct, using the same tools and investigating the same phenomena even if they sometimes vigorously disagree on the conclusions. The result is often that little progress is made. To avoid such an unseemly outcome, we thought it was imperative to include speakers and writers who do not talk to each other all the time, who do not hang out in the same social circles, who do not publish (or try to) in the same publications.

The result is the book you are holding in your hands. Most of the contributors—who include quite a number of invited authors who were unable to speak at the conference—are economists of one sort or another. However, macroeconomists, financial economists, business economists, labor economists, and public policy economists have less in common than you might think. True, they all accept the basic tenets of economic analysis: the importance of trade-offs, the power of incentives, and the tendency of markets to seek equilibrium. But in their specific applications, these subspecies of economists might as well be in different worlds.

By bringing them all together in a large room—and by including written contributions from those who couldn't come—we were able to build a kind of mini-university of productivity studies. The word *university* comes from the Latin for “whole” (as in “universe”). It implies that you can only see the whole

picture by convening scholars from widely separated disciplines. You can have a college of medicine or a college of commerce, but a university has to cover it all.

In medieval universities, which were the first of their kind, the disciplines were arithmetic, geometry, music, and astronomy—the “four ways,” or *quadrivium*. Also included were grammar, logic, and rhetoric—the “three ways,” or *trivium*. (Trivial, they’re not, although somebody must have thought so or we would not have inherited that meaning of the word.) In the present effort, we’re combining the various aspects of economics mentioned earlier, plus history, governance, political philosophy, and journalism.

A Precedent for Interdisciplinary Conferences

The *Life-Cycle Investing and Saving* conferences, organized by Zvi Bodie of Boston University and supported by the CFA Institute Research Foundation, likewise brought together people from all walks of life: “professors of finance, insurance, and risk management; investment managers . . . , actuaries . . . , lawmakers, lawyers, regulators, and accountants . . . [, and] pension plan executives—a broad category that includes corporate managers, government officials, [and] labor union representatives.”⁵ Diversity and inclusion are not new ideas to us.

Neither is having a global focus. While many of our speakers and article contributors concentrate on the United States—for that is where the bulk of the research has been produced—we’ve made a positive effort to broaden our reach to the world. One of our authors is Finnish, one Turkish, one Colombian, one Canadian, one Austrian, one French. We would have loved to include speakers from China, India, and Africa, but they were unavailable. In some ways, the developing world is now the technological frontier, where innovations first percolate to the surface; we’d like to cover that phenomenon in future work.

⁵Zvi Bodie, Dennis W. McLeavey, and Laurence B. Siegel, eds., *The Future of Life-Cycle Saving and Investing*, 2nd ed. (Charlottesville, VA: CFA Institute Research Foundation, 2008), <https://www.cfainstitute.org/en/research/foundation/2008/future-of-life-cycle-saving-investing-second-edition>.

Zvi Bodie, Laurence B. Siegel, and Rodney N. Sullivan, *The Future of Life-Cycle Saving and Investing: The Retirement Phase* (Charlottesville, VA: CFA Institute Research Foundation, 2010), <https://www.cfainstitute.org/en/research/foundation/2009/future-of-life-cycle-saving-investing-retirement-phase>.

Zvi Bodie, Laurence B. Siegel, and Lisa Stanton, *Life-Cycle Investing: Financial Education and Consumer Protection* (Charlottesville, VA: CFA Institute Research Foundation, 2012; corrected ed., January 2013), <https://www.cfainstitute.org/en/research/foundation/2012/life-cycle-investing-financial-education-and-consumer-protection-corrected-january-2013>.

Distinguished Contributors

Our “opening statement” is from the Columbia University professor Edmund Phelps, winner of the 2006 Nobel Memorial Prize in Economic Sciences (hereafter “the Nobel Prize,” clarifying my view on the silly and tiresome argument about whether it’s a real Nobel). Phelps’s role in setting the stage for the entire proceeding, by giving the keynote address, was invaluable. But we did not stop there in trying to get top-name authors to contribute to this monograph. We reached beyond the list of conference speakers to publish articles by authors we especially respect who could not attend the conference.

Among these are Deirdre McCloskey, to whom we gave the “last word”; she is the philosopher-economist who, remarkably, once held appointments in six academic departments at the same time: economics, history, English, communication, philosophy, and classics. (In the last two, she was an adjunct.) Martin Feldstein, advisor to presidents, writes on economic measurement. And Walter Russell Mead, the foreign policy scholar, writes on “infostructure,” the kind of infrastructure investment he thinks will most successfully spur communities to greater prosperity.

In his Preface, David Adler, who organized the conference, presents a full road map through this book. I just wanted to mention some highlights.

Conclusion

The CFA Institute Research Foundation has, by and large, published books (called monographs) that explore various aspects of investment finance. In recent years, we have broadened our ambit somewhat to include worthy contributions on related topics that might escape the notice of mainstream finance publications. One of these was the *Financial Market History* monograph, like the present volume based on a conference and drawing on a rich variety of people and ideas.⁶ We were especially proud of that effort, and we are just as delighted to be able to present this one.

Here, in the hope of doing some good and helping to foster not only discussion about productivity growth but actual change, we take another step away from our traditional area of focus. We are exceptionally pleased to disseminate *The Productivity Puzzle: Restoring Economic Dynamism* and hope that it spurs its readers to reinvigorate the economic landscape through creative entrepreneurship and thoughtful business and investment decisions worldwide.

⁶David Chambers and Elroy Dimson, eds., *Financial Market History: Reflections on the Past for Investors Today* (Charlottesville, VA: CFA Institute Research Foundation, 2016), <https://www.cfainstitute.org/en/research/foundation/2016/financial-market-history>.

Opening Statement

Right Values and Wrong Ones for an Economy of Dynamism

Edmund Phelps

Director of the Center on Capitalism and Society, Columbia University

The backdrop of this meeting is a welter of disturbing economic statistics dating back to the late 1960s or early 1970s: Foremost is the stagnation of productivity.⁷ This “stagnation” led to *depressed rates of return to investment, stagnant wage rates*—indeed, somewhat stagnant incomes generally—*bloated levels of household wealth relative to wages* as saving did not stop when wage income stopped growing, *increased social insurance outlays relative to wage rates*, an *explosion of public debt* and—unsurprisingly in light of all this—*more and more prime-age men staying outside the labor force*.⁸

True, unemployment rates and investment are looking good. Some of the recent horrors have abated or disappeared with the present *boom*, just as they did during the internet boom. But booms are *temporary*.

From *my* perspective, the problem in the West is that the *lead* countries are suffering from a loss of the old *mojo*—a loss of their former drive. There has been a significant *loss of the modernist spirit*, which stirred powerfully in late-Renaissance Italy and Germany (late 15th and 16th century)—think of Pico della Mirandola, Luther, and Cellini—and reached a “critical mass” in the 19th century: first Britain and America, later Germany and France.

I think of this “modernist spirit” as composed of several *human values*. One is *individualism*—thinking for yourself, willing to break from convention, and a Dickensian desire to “take control of your life.” The second is *vitalism*—having the Shakespearian courage to act, the Nietzschean will to surmount obstacles and, in Lincoln’s words, a “rage” for the new. The last is *expressionism*—exercising the imagination that Hume speaks of, exploring or experimenting, thus voyaging into the unknown. Modern people have a fascination for uncertainty.

Reprinted from the New York Museum of American Finance, 28 November 2017, with the permission of the publisher, Center on Capitalism and Society.

⁷Let me say to noneconomists that for economists, productivity means output per unit of labor or output per unit of capital or some average of the two, later called total factor productivity.

⁸One could add the weakness of the dollar—with the notable exception of the Clinton years.

It was the ultimate birth of this modern spirit—in the nations fortunate to have it—that fueled what I call *dynamism*, by which I mean a *desire* and a *capacity* to introduce welcome improvements. This dynamism sparked the *explosions of innovation* in the 19th century, and this innovation became the engine fueling business investment. (Of course, even a person or a nation possessing the utmost dynamism may *fail* to achieve the innovation they sought.)

How do we know these things? There is cross-country evidence from data on countries in the OECD that the degree of modernism, as measured by various attitudes and beliefs, is strongly correlated with the degree of economic performance, as measured by job satisfaction, labor force participation, etc.

It will not surprise you that I believe it is a *substantial loss* of that modernist spirit, thus some loss of dynamism, that has largely led to *much lower rates of innovation* starting around the *late 1960s*. (The years of the internet boom were a welcome respite, of course, and the Bush–Greenspan housing boom was not much of a respite and not welcome.) The losses of innovation appear to be located in the traditional industries. The innovation we *do* have appears to be concentrated in new, high-tech industries—in fact, just a few celebrated corporations, and recently, even they seem to have lost some of their impact on productivity.

Furthermore, in my book *Mass Flourishing*, I provide some evidence that innovation in the 19th century was also *pervasive*—in all or most industries—and *inclusive*—from the grassroots of society on up. Much, perhaps *most*, of the contribution by innovation to economic growth can be laid to the new ideas of *ordinary people* engaged in business life. The work they did every day led them to conceive of some better or different methods in farms, factories, and offices—though they must have been aware that commercial *success* was uncertain.

Now, it also appears to me, there is a *dread* of “Knightian” uncertainty (named after Frank Knight⁹—Keynes also introduced the concept).¹⁰ People came to be uncomfortable with the directionlessness that modernist values injected into the economy. The *loss* of their former fascination with voyaging into the unknown—which is an element of *expressionism*—is one of the causes of the serious loss of dynamism, thus a serious loss of innovation. (I did not say *disappearance* of innovation, only a serious *loss* of it.)

Some of us believe there has also been a decline of *individualism*. Where are the Horatio Alger stories? Where are the young people asking Horace Greeley in what direction to go? I am shocked that young people tell opinion surveys that they want to remain in their hometown, live close to their

⁹Frank H. Knight, *Risk, Uncertainty, and Profit* (Boston: Houghton Mifflin, 1921).

¹⁰John Maynard Keynes, *A Treatise on Probability* (London: Macmillan, 1921).

friends, or even continue to live at home!¹¹ This is a portrait of America that is almost unrecognizable to me. Certainly it is not the nation that Norman Rockwell painted and Willa Cather wrote about.

How about *vitalism*? Are Americans still OK on that score? I am not sure. I wonder, are Americans still do-ers? Do they love to compete as much as in the decades from, say, the 1850s up to the mid-1960s?

Or are they still the couch-potatoes that was once said about them? Are they fixed on all the *tweets* coming in by the hour?

There are other hypotheses on the causes of this stagnation, most of them a reversion to the tenets of corporatism: There is the rise of the “money culture,” as John Dewey called it. There is also the strange love affair of most Americans and Europeans (including Brits) with houses—Rome and New York are rare exceptions—which is another kind of materialism.

The flagrant *short-termism* of corporate heads and our representatives in legislatures—witness the tax cuts proposed in Washington—is another hypothesis. Answering a query from Larry Summers, I looked into what has happened to the steepness of the yield curve since the earliest period to recent periods. The trend has been up. In the period 1925–1932, the average 10-year rate was only 0.05 points above the average 3-month rate. In the period 1994–1996, it was 1.93, in 2003–2005 and in 2016–17 it was 1.51.¹² These observations are consistent with the hypothesis that asset managers and clients are more averse to long-term assets, with their relatively high element of uncertainty, than they were in the span of *normal* years in the interwar period. However, the hypothesized rise of short-termism is *not* outside my framework of *modernist values*. It looks to me like a loss of *vitalism*.

The emergence of *abusive use of patents* and *protectionist regulations* is yet another compelling hypothesis. The problems are too well-known to be set out here. I would only make the point that an economy needs some basic patent protection and some basic regulations; however, a forest of regulation and patents makes it burdensome for individuals to start new companies and presents legal hazards to employees and managers inside existing companies who would have liked to try out new methods or policies. Why has society allowed these governmental abuses to arise? In part, my answer is that much of the citizenry have lost their allegiance to modernist values.

¹¹See Brad Tuttle, “Being 30 and Living with Your Parents Isn’t Lame—It’s Awesome,” *Time*, 20 March 2012.

¹²Thomas S. Coleman, Lawrence Fisher, and Roger G. Ibbotson, *Historical U.S. Treasury Yield Curves* (New York: Moody Investor’s Services, 1993); for later data see the resource center on the US Department of the Treasury’s website.

Finally, politicians have taken ad hoc measures that directly *block competition from new ideas*. The entry of startup firms is impeded through a variety of actions—from tariffs and quotas to outright aid to incumbents—to save established companies from losing market share. Furthermore, when incumbents become safe from firms with new ideas, they can afford to cut back whatever defensive innovation they might have done. All this represents a serious rejection of individualism in favor of collective action.

So, we are faced with a revolt against the modern values that once drove massive innovation in the lead economies of the West and a rise of postmodern values that have gradually pointed society toward other ways of life. We will not be able to regain the dynamism of old unless we restore the modernist values and reject the postmodern ones.

History and Theory

A Blind Spot in Today's Macroeconomics?

Claudio Borio

Head of the Monetary and Economic Department, Bank for International Settlements

There is a blind spot in today's macroeconomics, and we have got so used to it that we hardly notice it. It is the idea that, for all intents and purposes, when making sense of first-order macroeconomic outcomes we can treat the economy as if its output were a single good produced by a single firm. To be sure, economists have worked hard to accommodate variety in goods and services at various levels of aggregation. Moreover, just to mention two, the distinctions between tradeables and non-tradeables or, in some intellectual strands, between consumption and investment goods have a long and distinguished history. But much of the academic and policy debate among macroeconomists hardly goes beyond that, if at all.

The presumption that, as a first approximation, macroeconomics can treat the economy as if it produced a single good through a single firm has important implications. It implies that aggregate demand shortfalls, economic fluctuations and the longer-term evolution of productivity can be properly understood without reference to intersectoral and intrasectoral developments. That is, it implies that whether an economy produces more of one good rather than another or, indeed, whether one firm is more efficient than another in producing the same good are matters that can be safely ignored when examining macroeconomic outcomes. In other words, issues concerned with resource misallocations do not shed much light on the macroeconomy.

In my remarks today, I would like to suggest that the link between resource misallocations and macroeconomic outcomes may well be tighter than we think. Ignoring it points to a kind of blind spot in today's macroeconomics. It would thus be desirable to bridge the gap, investigate the nexus further and explore its policy implications. Today's conference is a welcome sign that the intellectual mood may be changing.

As an illustration, I will address this question from one specific angle: the role of finance in macroeconomics. As we now know, the Great Financial Crisis (GFC) has put paid to the notion that finance is simply a veil of no consequence for the macroeconomy—another firmly and widely held notion that has proved inadequate. I will first suggest, based on some recent empirical

Reprinted with the permission of the publisher, Bank for International Settlements. Originally published 10–11 January 2018.

work, that the resource misallocations induced by large financial expansions and contractions (financial cycles) can cause material and long-lasting damage to productivity growth. I will then raise questions about the possible link between interest rates, resource misallocations and productivity. Here I will highlight the interaction between interest rates and the financial cycle and will also present some intriguing empirical regularities between the growing incidence of “zombie” firms in an economy and declining interest rates. I will finally draw some implications for further analysis and policy.

The Financial Cycle–Productivity Nexus

The GFC has hammered home the message that financial cycles can cause huge economic damage.¹³ As I like to stress, macroeconomics without the financial cycle is very much like Hamlet without the Prince (Borio 2014). The self-reinforcing interaction between credit, risk-taking and asset prices, especially property prices, can lead to self-sustained expansions and contractions that, when sufficiently large, can produce deep recessions, shallow recoveries and persistently lower growth, leaving long-lasting scars on the economic tissue. This is so especially when banking crises occur.¹⁴

In seeking to explain these stylised facts, the profession has focused on the demand side and, moreover, has tended to treat the economy as if it produced a single good. In other words, it has focused on the Okun (or output) gap (Okun 1962) as if its composition did not matter.¹⁵ This is natural in some

¹³See, for instance, Drehmann, Borio, and Tsatsaronis (2012), Claessens, Kose, and Terrones (2011), Aikman, Haldane, and Nelson (2015), Jordà, Schularick, and Taylor (2013), Juselius and Drehmann (2015), and using a different terminology, Reinhart and Rogoff (2009). For earlier studies of the financial cycle, see Borio, Furfine, and Lowe (2001) or Borio and Lowe (2002) and for a recent survey of the literature, Claessens and Kose (2017). Drehmann et al. (2012) also include references to previous work, as the notion of the financial cycle in fact predates that of the business cycle. This perspective contrasts starkly with that put forward by Lucas (2003).

¹⁴See the Basal Committee on Banking Supervision (2010) survey and, in particular, Cerra and Saxena (2008); for a more recent estimate, see Ball (2014). For more general evidence that bigger financial booms are associated with deeper recessions and longer recoveries after a financial bust, see Claessens et al. (2011), Drehmann et al. (2012) and Jordà et al. (2013). For the costs of household credit booms, see in particular Mian, Sufi, and Verner (2015) and Drehmann, Juselius, and Korinek (2017).

¹⁵This is not to deny the existence of models that consider explicitly one additional sector, such as housing. For instance, Iacoviello and Neri (2010) provide a model with housing and non-housing production where housing prices affect household borrowing capacity. More recently, Kaplan, Mitman, and Violante (2017), using a model that allows for house renting, show that the boom-bust cycle in house prices explains half the fluctuations in non-durable consumption in the United States.

respects. What's more, there is no question that the financial cycle-induced collapse in expenditures is the main factor behind the damage we have seen.

But is this all? Might not resource misallocations also have played a role? Might they not have interacted closely with macroeconomic outcomes through their influence on both productive potential and their link with aggregate demand? This is indeed what we find in some recent work (Borio et al. 2015).

We proceed as follows. First, we decompose the evolution of labour productivity growth into a component that is common to all economic sectors and one that results from labour shifts across sectors—purely an identity.¹⁶ We study labour shifts only because of data limitations: capital could be even more important, as sector-specific capital overhangs can have even longer-lasting effects. Think of overbuilding, for instance. Second, we explore how far each component is explained by measures of a credit boom, controlling for the influence of other factors. Finally, we examine how well the behaviour of the two productivity components during credit booms predicts the behaviour of productivity during subsequent recessions and their aftermath.¹⁷ Here, too, we control for the influence of other factors and examine, in particular, how the evolution of productivity depends on whether a banking crisis occurs or not. We do all this in a sample of over 21 advanced economies over the period 1969–2013.

We come up with two key findings.

First, credit booms tend to undermine productivity growth as they occur. For a typical credit boom, a loss of just over a quarter of a percentage point per year is a kind of lower bound (**Figure 1**, Panel A). The key mechanism is the credit boom's impact on labour shifts towards lower productivity growth sectors, notably a temporarily bloated construction sector. That is, there is an economically and statistically significant relationship between credit expansion and the allocation component of productivity growth (compare Panel A with Panel B of **Figure 2**). This mechanism accounts for slightly less than two thirds of the overall impact on productivity growth (**Figure 1**, Panel A, blue portion). In other words, not only do credit booms undermine productivity growth, as already found by Cecchetti and Kharroubi (2015), but they do so mainly by inducing shifts of resources into lower productivity growth sectors.

Second, the subsequent impact of the labour reallocations that occur during a financial boom is much larger if a banking crisis follows. The average

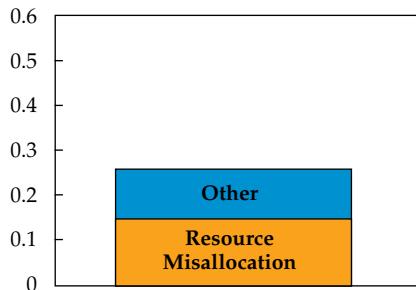
¹⁶We borrow the decomposition from Olley and Pakes (1996), although they apply it across firms within a given sector.

¹⁷This is a local-linear projection-type regression in the spirit of Jordà et al. (2013).

Figure 1. Financial Booms Sap Productivity by Misallocating Resources

A. Annual Cost during a Typical Boom

Percentage Points



B. Annual Cost over a Five-Year Window Post-Crisis

Percentage Points



Notes: Estimates calculated over the period 1969–2013 for 21 advanced economies. Resource misallocation = annual impact on productivity growth of labour shifts into less productive sectors during a five-year credit boom and over the period shown. Other = annual impact in the absence of reallocations during the boom.

Source: Borio et al. (2015).

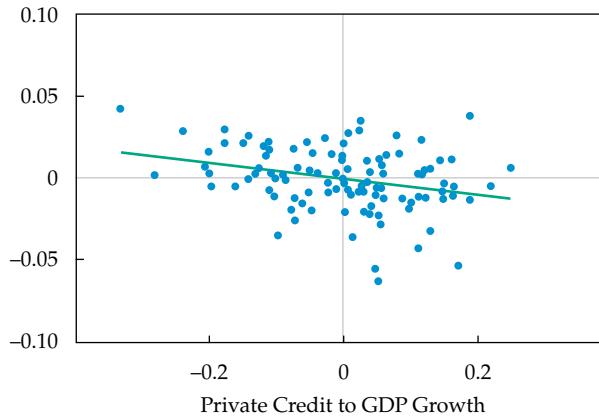
loss per year in the five years after a crisis is more than twice that during the boom, around half a percentage point per year (Figure 1, Panel B). Indeed, as shown in the simulation presented in **Figure 3**, the impact of productivity growth in that case is very long-lasting. The reallocations cast a long shadow.

The overall effects can be sizeable. Taking, say, a (synthetic) five-year credit boom and five postcrisis years together, the cumulative shortfall in productivity growth would amount to some 6 percentage points. Put differently, for the period 2008–2013, we are talking about a loss of some 0.6 percentage points per year for the advanced economies that saw booms and crises. This is roughly equal to their actual average productivity growth

Figure 2. Financial Booms and Productivity Growth Components (computed over five-year windows and taken as deviations from country and period means)

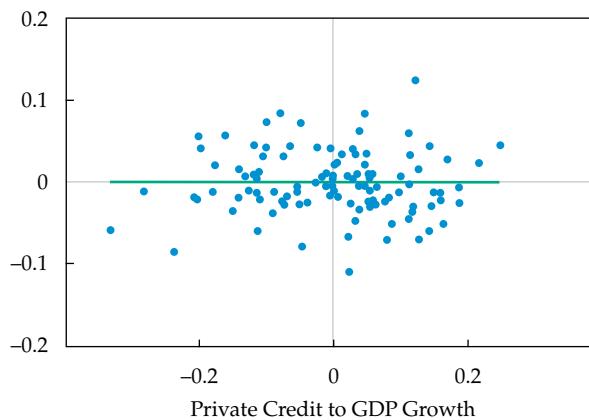
A. Credit Booms and the Allocation Component

Allocation Component
Contribution to Productivity Growth



B. Credit Booms and the Common Component

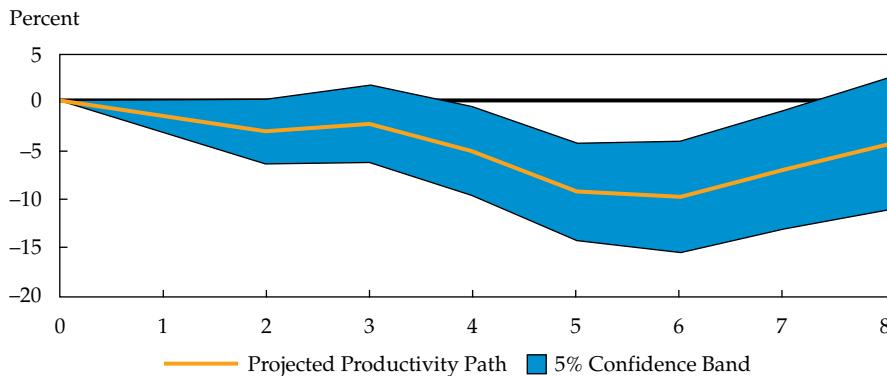
Common Component
Contribution to Productivity Growth



Notes: The panels plot the growth rate in private credit to GDP against the allocation and common components of labour productivity growth, respectively; both variables are taken as deviations from country and period means. The sample includes 21 economies and six periods of five years (1979–1984; 1984–1989; 1989–1994; 1994–1999; 1999–2004; 2004–2009).

Source: Borio et al. (2015).

Figure 3. Productivity Stagnates after a Financial Crisis due to Previous Labour Misallocations



Notes: This simulation is based on local projection regressions of the percentage deviation of labour productivity from the recession year. The independent variables include the allocation and the common components of productivity growth over the three-year period prior to the start of the recession. The orange line shows the projection of labour productivity conditional on the occurrence of a financial crisis and a positive allocation component contribution of 0.85 percentage points over three years prior to the recession (first quartile of the distribution of the allocation component contribution). The blue area around the orange line represents the 5% confidence interval around the projected productivity path.

Source: Borio et al. (2015).

during the same window. Now, the point is not to take these magnitudes at face value, but to note that these factors are material and should receive much more attention.

How could one explain the much larger impact of the misallocation of resources during the boom when a banking crisis follows? More research is needed, but a reasonable conjecture is that sectors that have expanded too much during the boom have to contract at some later stage—this is what allows us to talk about “misallocations” in the first place as opposed to mere reallocations. In this vein, the larger costs in the wake of a banking crisis may reflect, at least in part, how overindebtedness and a broken banking system hinder the required adjustment. For instance, if households are underwater, with mortgage debt exceeding the value of their house, they will find it harder to relocate to take advantage of job opportunities.¹⁸ More to the point, banks with impaired balance sheets and high non-performing loans have strong incentives not to recognise losses and to misallocate credit: they will tend to

¹⁸See, for instance, Demyanyk et al. (2017) for the United States and Henley (1998) for the United Kingdom.

keep the spigots open for weaker borrowers (“evergreening”) while curtailing or increasing the cost of credit to healthier ones, which can afford to pay. Evidence confirms this.¹⁹

The analysis also enriches our understanding of how the productive capacity of the economy can be persistently weakened. Keeping with the presumption that the economy can be treated as if it produced a single good, macroeconomists have long recognised that persistent shortfalls in aggregate demand can sap supply: the unemployed lose their skills and a lack of investment undermines future productive potential, not least by slowing down the adoption of new technologies (Hall 2014; Reifschneider, Wascher, and Wilcox 2015; Anzoategui et al. 2016). But the structure of production matters too. The sectors that overexpand then need to contract towards a more sustainable size, in the process complicating the economy’s adjustment to expansions and contractions in aggregate demand. Here the core of the problem is not the generalised shortfall of demand but the misallocation of resources across sectors; in fact, the abundance of aggregate demand during the boom is what helps promote the misallocation in the first place. This also means that the cure has to address the misallocation itself. This form of “hysteresis” or history-dependence definitely requires further study.

And what is true across sectors may also apply within sectors. There is indeed a growing body of work exploring this dimension, some of which is on display at this conference (e.g., Linarello, Petrella, and Sette 2017).

From an analytical perspective, working at firm, rather than sector, level has advantages. Number of observations aside, a key one is that it is easier to control for “demand” factors: all firms in the same sector can more plausibly be assumed to face a similar demand for their output. This makes it easier to identify “misallocations”²⁰ and to tell them apart from differences that reflect the exposure to different demand (e.g., Gopinath et al. 2017). For example, lower-productivity sectors may expand simply because as the economy grows richer, the demand for their output rises faster than that for other sectors, rather than because of a long-lasting imbalance. But this should not discourage the researcher from studying sectoral misallocations. From a macroeconomic perspective, they may be even more important.

¹⁹See, for instance, Caballero, Hoshi, and Kashyap (2008), for Japan, Acharya et al. (2016) for the European experience, and Albertazzi and Marchetti (2010) and Schiavardi, Sette, and Tabellini (2017) for that of Italy.

²⁰The misallocations so identified may be of a different kind to the intersectoral ones mentioned above, which have an inherent intertemporal character (temporary demand-driven overexpansion that at some point needs to be reversed).

The Interest Rate–Productivity Nexus

So much for the link between the financial cycle, resource misallocations, and productivity. Might there not be also a link between interest rates, in particular persistently low interest rates, and productivity?

The standard argument is that such a link may well exist, but with causality running from productivity growth to real (inflation-adjusted) interest rates. This takes root in the notion that, over long enough periods, the real economy evolves independently of monetary policy—in jargon, “money neutrality.” In that case, one can also presume that market interest rates converge to an equilibrium real interest rate (or natural rate) that depends exclusively on non-monetary factors (e.g., Bernanke 2005; Summers 2014; Bean, Broda, Ito, and Kroszner 2015). Under some auxiliary assumptions, productivity growth would be one such factor.

I have argued in detail elsewhere why this view may be overly simplistic and may play down too much the role of monetary factors (Borio 2017a). Granted, it is *a priori* reasonable to expect that productivity growth would influence real interest rates. But the empirical evidence indicates that, in general, the link between real interest rates and productivity has been rather tenuous historically (Hamilton et al. 2015; Lunsford and West 2017; Borio et al. 2017). And this evidence does not preclude the possibility that, under some conditions, the link may be present over horizons relevant for policy and that both real and nominal interest rates may matter too.

One way a link may exist, but with causality running from interest rates to productivity, is through the interaction between interest rates and the financial cycle. Here, the policy response is critical. Consider in turn the expansion and contraction phases of the cycle.

During the expansion phase, low interest rates, especially if persistent, are likely to increase the cycle’s amplitude and length. After all, one way in which monetary policy operates is precisely by boosting credit, asset prices, and risk-taking. Indeed, there is plenty of evidence to this effect.²¹ Moreover, the impact of low interest rates is unlikely to be uniform across the economy. Sectors naturally differ in their interest rate sensitivity. And so do firms within a given sector, depending on their need for external funds and ability to tap markets. For instance, the firms’ age, size, and collateral availability matter. To the extent that low interest rates boost financial booms and

²¹For evidence on the impact of monetary policy on credit and asset prices, see, e.g., Calza, Monacelli, and Stracca (2013), Bauer and Granziera (2017), Juselius et al. (2017), and Hofmann and Peersman (2017a, 2017b). For a description of the risk-taking channel and evidence, see, e.g., Borio and Zhu (2012), Adrian and Shin (2010), Jimenez et al. (2014), Dell’Ariccia, Laeven, and Suarez (2017), and Cecchetti, Mancini-Griffoli, and Narita (2017).

induce resource shifts into sectors such as construction or finance, they will also influence the evolution of productivity, especially if a banking crisis follows. Since financial cycles can be quite long—up to 16 to 20 years—and their impact on productivity growth quite persistent, thinking of changes in interest rates (monetary policy) as “neutral” is not helpful over relevant policy horizons.²²

During the financial contraction, persistently low interest rates can contribute to this outcome (Borio 2014). To be absolutely clear: low rates following a financial bust are welcome and necessary to stabilise the economy and prevent a downward spiral between the financial system and output. This is what the crisis management phase is all about. The question concerns the possible collateral damage of persistently and unusually low rates thereafter, when the priority is to repair balance sheets in the crisis resolution phase. Granted, low rates lighten borrowers’ heavy debt burden, especially when that debt is at variable rates or can be refinanced at no cost. But they may also slow down the necessary balance sheet repair.

There are at least a couple of reasons for this. Persistently low rates may interact with bank weakness to delay the resolution of underlying balance sheet problems. It is easier to carry bad loans when their opportunity cost is lower. And it is more difficult to discriminate across borrowers when interest rates are very low, delaying their balance sheet repair. Ultimately, unprofitable firms could survive for longer, crowding out resources for the rest (“zombie lending”).

While these mechanisms are quite plausible, specific empirical evidence is rather scant. As noted, most of the evidence relates to bank weakness as such rather than to the impact of persistently low interest rates per se. Clearly, distinguishing the two is not easy, as they would tend to coexist. There is also evidence that large-scale asset purchases have compressed credit risk premia (Gilchrist and Zakrjsek 2013; Rogers, Scotti, and Wright 2014; Altavilla, Carboni, and Motto 2015) and, especially when buying the corporate assets themselves, helped reduce risk differentiation (ECB 2017a, 2017b).²³ But the impact of this effect on resource allocation has not been quantified. More analysis would be welcome.

Could there be a more general relationship between the level of interest rates and the incidence of unprofitable firms that survive? My colleagues

²²See also Borio et al. (2017) for a more general empirical analysis finding evidence against money neutrality in this sense.

²³An example of this lower credit risk differentiation is that firms have found it cheaper to issue in euros and then swap into dollars than to issue in dollars directly, in the process putting pressure on the cross-currency basis; see Borio et al. (2016).

Ryan Banerjee and Boris Hofmann (2018) have begun to examine this possibility. They take as a starting point the definition of “zombie firms” that the OECD has employed in its excellent research in this area—firms that are at least 10 years old and whose profits (EBIT) are insufficient to cover interest payments (e.g., Adalet McGowan, Andrews, and Millot 2017). Then they further refine the definition by restricting zombie firms to be those with comparatively low expected future growth potential, i.e., those that also have below-median Tobin’s Q within a sector in a given year.²⁴ The more restrictive definition is intended to exclude the Teslas of this world—loss-making firms with strong future growth potential. They then examine a sample of nearly 32,000 publicly quoted firms from 14 OECD countries going as far back as 1980.²⁵

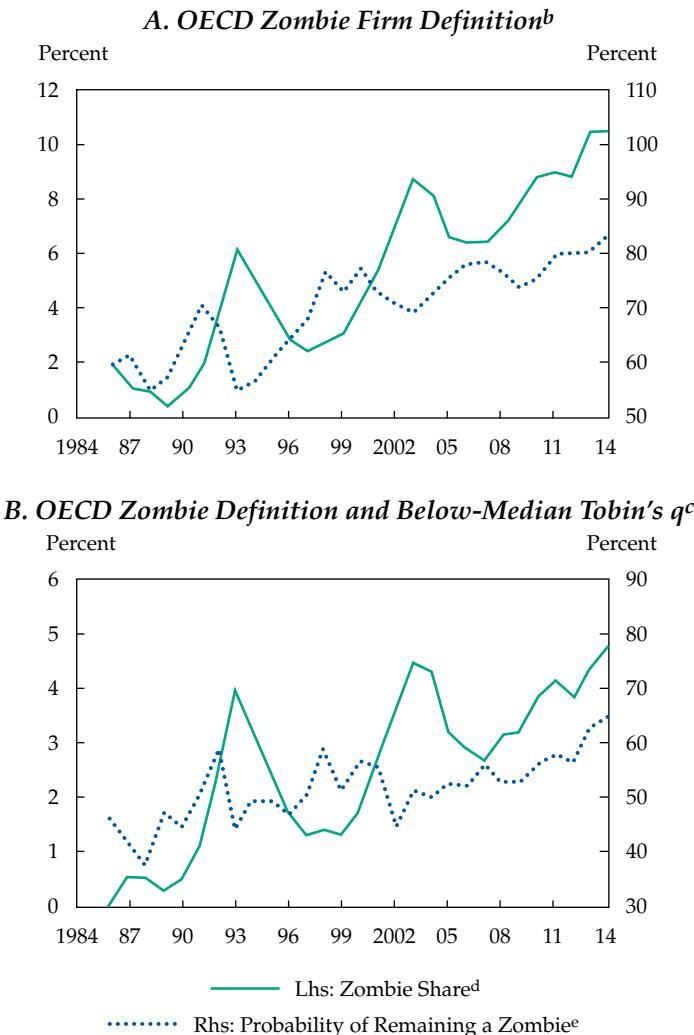
The first point to note is that zombies have been on the rise and survive—if I can use that term—for longer. Cyclical variations aside, the mean share of publicly quoted zombie firms across these economies has steadily trended up, from close to zero to above 10% under the OECD definition (**Figure 4**, Panel A), and up to 5% under the more restrictive one (Panel B). Furthermore, zombies remain in that state for longer (both panels). For instance, based on the narrower definition, in 1987 the probability of a zombie firm remaining a zombie in the following year was approximately 40%; by 2016 it had risen to 65%. That probability is even higher based on the OECD definition.

How are zombies today able to survive for longer than they did in the 1980s and 1990s? The answer at this stage can only be very preliminary. But one possibility is that, since the early 2000s, they seem to face less pressure to reduce debt. Regression estimates suggest that pre-2000 zombies cut debt at a rate of over 3% of total assets per year relative to non-zombie firms; but post-2000, the two groups are practically indistinguishable (**Figure 5**). At the same time, interest payments on their debt declined even in relative terms (same graph). And this occurred even as zombies’ coverage ratio also worsened over time (not shown).

The counterpart to this ability to avoid reducing debt is that zombie firms have been locking in more resources, hindering the reallocation process.

²⁴This refinement of the zombie firm definition also produces stronger zombie congestion effects, whereby a greater share of zombies in a sector reduces investment and employment of non-zombie firms.

²⁵Focusing on publicly quoted firms has two main advantages. First, the longer time span of data on these firms allows analysis over several business cycles. Second, it is possible to take into account the perceived future growth potential as reflected in equity prices. However, a significant drawback is that publicly quoted firms are only a limited subset of the universe of firms in an economy.

Figure 4. Zombie Firms on the Rise and Surviving for Longer^a

^aSample includes listed non-financial firms in Australia, Belgium, Canada, Denmark, France, Germany, Italy, Japan, the Netherlands, Spain, Sweden, Switzerland, the United Kingdom, and the United States.

^bZombie firm defined as a firm whose EBIT is below interest payments and is over 10 years old.

^cZombie firm defined as a firm whose EBIT is below interest payments, is over 10 years old, and has below-median Tobin's Q in its two-digit SIC sector in a given year.

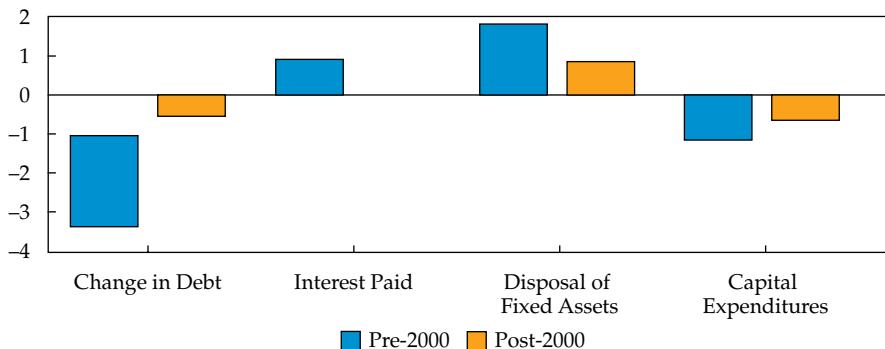
^dRatio of the number of zombie firms to all listed firms.

^eProbability of a firm remaining a zombie in the following year, conditional on it being a zombie in the current year.

Source: Banerjee and Hofmann (2018).

Figure 5. Zombies' Behaviour Relative to Non-Zombies: Over Time, Less Pressure to Cut Debt^a

Share of Total Assets



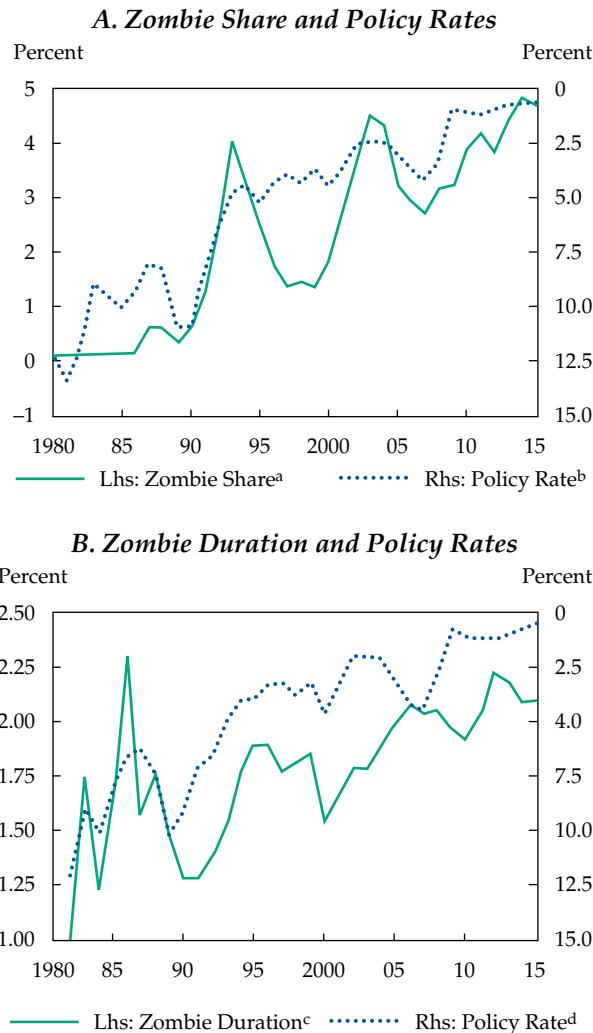
^aEstimates of β_1 and β_2 from the regression $y_{i,t} = \beta_1 D(\text{zombie}_{\text{pre-2000}}) + \beta_2 D(\text{zombie}_{\text{post-2000}}) + \gamma \text{Controls}_{i,t} + \alpha_{\text{sector},t} + \delta_{\text{country},t} + \varepsilon_{i,t}$, where $D(\text{zombie}_i)$ is a dummy variable indicating whether firm i is classified as a zombie in period t . Control variables: ratio of fixed assets to total assets, industry cash flow volatility, market-to-book value, logarithm of total assets in 2010 US dollars, ratio of capital expenditures to total assets, ratio of R&D to sales, dummy variable indicating whether the firm pays a dividend. $\alpha_{\text{sector},t}$, $\delta_{\text{country},t}$ are sector-year and country-year dummy variables, respectively. Source: Banerjee and Hofmann (2018).

Relative to their more profitable peers, they have slowed down asset disposals and refrained from cutting capital expenditure (Figure 5).

But why should they have been better able to bear debt? What's intriguing and striking is the close relationship between the incidence of zombies and the decline in nominal interest rates (Figure 6, Panel A; the axis for interest rates is inverted). The relationship is remarkably tight.²⁶ This is true even if one excludes cyclical variations, in which, say, lower rates may reflect weak aggregate conditions. Moreover, the relationship does not derive so much from temporarily unprofitable firms. The length of time firms remain unprofitable increases as interest rates decline (Panel B).

Now, the relationship could be purely coincidental. Possible factors, unrelated to interest rates as such, might help explain the observed relationship. One other possibility is reverse causality: weaker profitability, as productivity and economic activity decline in the aggregate, would tend to induce central banks to ease policy and reduce interest rates. This no doubt helps explain the cyclical variations but is less compelling as an explanation of the trend and the ratcheting up. Still, the relationship may also point to a deeper link between interest rates and zombie firms. As argued above, under

²⁶This is the case regardless of which of the two definitions of zombie firms is used.

Figure 6. Zombies Rise and Survive for Longer as Nominal Interest Rates Decline

^aMean of country zombie shares; narrower definition.

^bMean of country (nominal) policy rates.

^cMean of zombie firm duration in years; narrower definition.

^dMean of (nominal) policy rates.

Source: Banerjee and Hofmann (2018).

some conditions low rates may generate long-lasting damage to productivity, including by amplifying the financial cycle, thereby contributing to the emergence of zombie firms. In turn, persistently low rates may also encourage lenders to be more forgiving, especially as they hunt for yield and/or find

the opportunity cost of not pulling the plug lower, thereby allowing zombies to survive for longer.

At this stage, we simply do not know enough. But a better understanding of the link would have significant implications for our understanding of what factors can drive resource misallocations and for policy. The OECD, for instance, has highlighted the drag on aggregate productivity that zombies can induce, both directly and indirectly (Adalet McGowan et al. 2017). What's more, regardless of the reasons, the higher incidence of zombie firms makes the economy more vulnerable to increases in interest rates to more normal levels—an aspect of what I have elsewhere described as a debt trap (e.g., Borio 2017b). All this raises difficult policy issues, ranging from appropriate targeted measures to broader structural and macroeconomic policies. Non-trivial trade-offs exist, especially in the short run.²⁷ Undoubtedly, the stylised fact deserves further study.²⁸

Conclusion

Let me conclude by highlighting the key takeaways of my remarks for analytics and policy.

I believe we need to go beyond the stark distinction between resource allocation and aggregate macroeconomic outcomes often implicit in current analysis and debates—a kind of blind spot in today's macroeconomics. There is a lot to be learned from studying their interaction as opposed to stressing their independence. I have illustrated this with a focus on the long-neglected link between finance and macroeconomic fluctuations. The financial cycle can cause first-order and long-lasting damage to productivity growth through its

²⁷For instance, it is generally recognised that it is always necessary to assess the firms' underlying prospects and then to proceed in an orderly way, based on that information and taking into account macroeconomic conditions. There are also trade-offs between different types of insolvency arrangements and adjustment speeds. Temporary support can play a useful role as part of a broader systematic strategy. Moreover, there are important issues about how to facilitate the redeployment of the resources that are released and limit the associated costs. Clearly, the longer firms remain in zombie status, the higher the costs relative to the benefits. And the costs are larger if banks are weak and not properly restructured, by recognising losses and recapitalising. For a detailed discussion of policies, see Andrews, Adelet McGowan, and Millot (2017) and, for one of some of the trade-offs involved, Haldane (2017).

²⁸Another interesting issue concerns cross-border effects, which may work on both borrowers and lenders. For instance, a large presence of zombie firms in one economy may cause congestion externalities on firm investment and employment in other economies, through exports or the operation of subsidiaries of multinational companies. Similarly, multinational banks may help transmit spillovers. And low interest rates in one economy may ease funding conditions elsewhere through a variety of mechanisms.

impact on resource misallocations. And we need to understand much better also the possible link between interest rates and such misallocations.

Policy, too, needs to be much better aware of these interactions. Some lessons are well understood, if not always put into practice. For instance, one such example is the need to tackle balance sheet repair head-on following a banking crisis so as to lay the basis for a strong and sustainable recovery (e.g., Borio, Vale, and von Peter 2010; Bech, Gambacorta, and Kharroubi 2014). Such a strategy is also important to relieve pressure on monetary policy. Doing so, however, has proved quite difficult in some jurisdictions following the GFC (e.g., Enria 2012; Borio 2016). Other aspects need to be better incorporated into policy considerations. The impact of persistently low rates is one of them. How well all of this is done may well hold one of the keys to the resolution of the current policy challenges.

I would like to thank Ryan Banerjee, Boris Hofmann, Enisse Kharroubi, and Fabrizio Zampolli for their help in the preparation of these remarks and Stijn Claessens, Piti Disyatat, David Laidler, Robert McCauley, Hyun Song Shin, and Kostas Tsatsaronis for their valuable comments.

References

Acharya, V., T. Eisert, C. Eufinger, and C. Hirsch. 2016. “Whatever It Takes: The Real Effects of Unconventional Monetary Policy.” Mimeo.

Adalet McGowan, M., D. Andrews, and V. Millot. 2017. “The Walking Dead? Zombie Firms and Productivity Performance in OECD Countries.” OECD Economics Department Working Paper 1372.

Adrian, T., and H. Shin. 2010. “Financial Intermediaries and Monetary Economics.” In *Handbook of Monetary Economics 3A*, edited by B. Friedman and M. Woodford, 601–650. Elsevier.

Aikman, D., A. Haldane, and B. Nelson. 2015. “Curbing the Credit Cycle.” *Economic Journal* 125 (585): 1072–1109.

Albertazzi, U., and D. Marchetti. 2010. “Credit Supply, Flight to Quality and Evergreening: An Analysis of Bank-Firm Relationships after Lehman.” Bank of Italy Working Paper 756.

Altavilla, C., G. Carboni, and R. Motto. 2015. “Asset Purchase Programmes and Financial Markets: Lessons from the Euro Area.” ECB Working Paper 1864. <https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1864.en.pdf>.

Andrews, D., M. Adalet McGowan, and V. Millot. 2017. “Confronting the Zombies: Policies for Productivity Revival.” *OECD Economic Policy Paper* 21.

Anzoategui, D., D. Comin, M. Gertler, and J. Martinez. 2016. “Endogenous Technology Adoption and R&D as Sources of Business Cycle Persistence.” *NBER Working Paper* 22005.

Ball, L. 2014. “Long-Term Damage from the Great Recession in OECD Countries.” *European Journal of Economics and Economic Policies* 11 (2): 149–60.

Banerjee, R., and B. Hofmann. 2018. “Corporate Zombies: Anatomy and Life Cycle.” Mimeo.

Basel Committee on Banking Supervision. 2010. “An Assessment of the Long-Term Economic Impact of Stronger Capital and Liquidity Requirements.” Basel, Switzerland: Bank for International Settlements. <https://www.bis.org/publ/bcbs173.pdf>.

Bauer, G., and E. Granziera. 2017. “Monetary Policy, Private Debt, and Financial Stability Risks.” *International Journal of Central Banking* 13 (3): 337–73.

Bean, C., C. Broda, T. Ito, and R. Kroszner. 2015. “Low for Long? Causes and Consequences of Persistently Low Interest Rates.” *Geneva Reports on the World Economy* 17. International Center for Monetary and Banking Studies.

Bech, M., L. Gambacorta, and E. Kharroubi. 2014. “Monetary Policy in a Downturn: Are Financial Crises Special?” *International Finance* 17 (1): 99–119.

Bernanke, B. 2005. “The Global Saving Glut and the U.S. Current Account Deficit.” Sandridge Lecture, Virginia Association of Economists, Richmond, VA, 10 March. <https://www.federalreserve.gov/boarddocs/speeches/2005/200503102/>.

Borio, C. 2014. “The Financial Cycle and Macroeconomics: What Have We Learnt?” *Journal of Banking & Finance* 45 (August): 182–98.

———. 2016. “The Banking Industry: Struggling to Move on.” Keynote speech at the fifth EBA Research Workshop, London, 28–29 November. <https://www.bis.org/speeches/sp161128.htm>.

———. 2017a. “Through the Looking Glass.” OMFIF City Lecture, London, 22 September. <https://www.omfif.org/media/3449020/claudio-borio-through-the-looking-glass.pdf>.

———. 2017b. “Secular Stagnation or Financial Cycle Drag?” *Business Economics* 52 (2): 87–98.

Borio, C., P. Disyatat, M. Juselius, and P. Rungcharoenkitkul. 2017. "Why So Long for So Long? A Long-Term View of Real Interest Rates." BIS Working Paper 685. <https://www.bis.org/publ/work685.htm>.

Borio, C., C. Furfine, and P. Lowe. 2001. "Procyclicality of the Financial System and Financial Stability: Issues and Policy Options." BIS Background Paper for *BIS Paper 1* (March). <https://www.bis.org/publ/bppdf/bispap01a.pdf>.

Borio, C., E. Kharroubi, C. Uppu, and F. Zampolli. 2015. "Labour Reallocation and Productivity Dynamics: Financial Causes, Real Consequences." BIS Working Paper 534. <https://www.bis.org/publ/work534.htm>.

Borio, C., and P. Lowe. 2002. "Asset Prices, Financial and Monetary Stability: Exploring the Nexus." BIS Working Paper 114.

Borio, C., R. McCauley, P. McGuire, and V. Sushko. 2016. "Covered Interest Parity Lost: Understanding the Cross-Currency Basis." *BIS Quarterly Review* (September): 45–64. https://www.bis.org/publ/qtrpdf/r_qt1609e.htm.

Borio, C., B. Vale, and G. von Peter. 2010. "Resolving the Financial Crisis: Are We Heeding the Lessons from the Nordics?" BIS Working Paper 311.

Borio, C., and H. Zhu. 2012. "Capital Regulation, Risk-Taking, and Monetary Policy: A Missing Link in the Transmission Mechanism?" *Journal of Financial Stability* 8 (4): 236–51.

Caballero, R., T. Hoshi, and A. Kashyap. 2008. "Zombie Lending and Depressed Restructuring in Japan." *American Economic Review* 98 (5): 1943–77.

Calza, A., T. Monacelli, and L. Stracca. 2013. "Housing Finance and Monetary Policy." *Journal of the European Economic Association* 11 (supplement 1): 101–22.

Cecchetti, S., and E. Kharroubi. 2015. "Why Does Financial Sector Growth Crowd Out Real Economic Growth?" BIS Working Paper 490. <https://www.bis.org/publ/work490.htm>.

Cecchetti, S., T. Mancini-Griffoli, and M. Narita. 2017. "Does Prolonged Monetary Easing Increase Financial Vulnerability?" IMF Working Paper 2017/65.

Cerra, V., and S. Saxena. 2008. "Growth Dynamics: The Myth of Economic Recovery." *American Economic Review* 98 (1): 439–57.

Claessens, S., and M. Kose. 2017. "Macroeconomic Implications of Financial Imperfections: A Survey." Policy Research Working Papers, World Bank Group.

Claessens, S., M. Kose, and M. Terrones. 2011. “Financial Cycles: What? How? When?” IMF Working Paper 11/76. <https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Financial-Cycles-What-How-When-24775>.

Dell’Ariccia, G., L. Laeven, and G. Suarez. 2017. “Bank Leverage and Monetary Policy’s Risk-Taking Channel: Evidence from the United States.” *Journal of Finance* 72 (2): 613–54.

Demyanyk, Y., D. Hryshko, M. Luengo-Prado, and B. Sørensen. 2017. “Moving to a Job: The Role of Home Equity, Debt, and Access to Credit.” *American Economic Journal: Macroeconomics* 9 (2): 149–81.

Drehmann, M., C. Borio, and K. Tsatsaronis. 2012. “Characterising the Financial Cycle: Don’t Lose Sight of the Medium Term!” BIS Working Paper 380.

Drehmann, M., M. Juselius, and A. Korinek. 2017. “Accounting for Debt Service: The Painful Legacy of Credit Booms.” BIS Working Paper 645.

ECB (European Central Bank). 2017a. *Economic Bulletin*, issue 4.

———. 2017b. *Financial Stability Review* (November).

Enria, A. 2012. “Supervisory Policies and Bank Deleveraging: A European Perspective.” Speech given at the 21st Annual Hyman P. Minsky Conference on the State of the US and World Economies, 11 April.

Gilchrist, S., and E. Zakrajsek. 2013. “The Impact of the Federal Reserve’s Large-Scale Asset Purchase Programs on Corporate Credit Risk.” *Journal of Money, Credit and Banking* 45 (s2): 29–57.

Gopinath, G., S. Kalemli-Özcan, L. Karabarbounis and C. Villegas-Sánchez. 2017. “Capital Allocation and Productivity in South Europe.” *Quarterly Journal of Economics* 132 (4): 1915–67.

Haldane, A. 2017. “Productivity Puzzles.” Speech given at the London School of Economics, 20 March. <https://www.bankofengland.co.uk/speech/2017/productivity-puzzles>.

Hall, R. 2014. “Quantifying the Lasting Harm to the U.S. Economy from the Financial Crisis.” *NBER Macroeconomics Annual* 29 (1): 71–128.

Hamilton, J., E. Harris, J. Hatzis, and K. West. 2015. “The Equilibrium Real Funds Rate: Past, Present, and Future.” NBER Working Paper 21476.

Henley, A. 1998. “Residential Mobility, Housing Equity and the Labour Market.” *Economic Journal* 108 (447): 414–27.

Hofmann, B., and G. Peersman. 2017a. "Monetary Policy Transmission and Trade-Offs in the United States: Old and New." BIS Working Paper 649. <https://www.bis.org/publ/work649.htm>.

———. 2017b. "Is There a Debt Service Channel of Monetary Transmission?" *BIS Quarterly Review* (December): 23–37. https://www.bis.org/publ/qtrpdf/r_qt1712e.htm.

Iacoviello, M., and S. Neri. 2010. "Housing Market Spillovers: Evidence from an Estimated DSGE Model." *American Economic Journal: Macroeconomics* 2 (2): 125–64. <https://www.aeaweb.org/articles?id=10.1257/mac.2.2.125>.

Jiménez, G., S. Ongena, J.-L. Peydró, and J. Saurina. 2014. "Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say about the Effects of Monetary Policy on Credit Risk-Taking?" *Econometrica* 82 (2): 463–505.

Jordà, O., M. Schularick, and A. Taylor. 2013. "When Credit Bites Back." *Journal of Money, Credit and Banking* 45 (s2): 3–28.

Juselius, M., C. Borio, P. Disyatat, and M. Drehmann. 2017. "Monetary Policy, the Financial Cycle and Ultra-Low Interest Rates." *International Journal of Central Banking* 13 (3): 55–89.

Juselius, M., and M. Drehmann. 2015. "Leverage Dynamics and the Real Burden of Debt." BIS Working Paper 501. <https://www.bis.org/publ/work501.htm>.

Kaplan, G., K. Mitman, and G. Violante. 2017. "The Housing Boom and Bust: Model Meets Evidence." NBER Working Paper 23694.

Linarello, A., A. Petrella, and E. Sette. 2017. "Allocative Efficiency and Finance." Mimeo.

Lucas, R. 2003. "Macroeconomic Priorities." *American Economic Review* 93 (1): 1–14.

Lunsford, K., and K. West. 2017. "Some Evidence on Secular Drivers of U.S. Safe Real Rates." FRB of Cleveland Working Paper 17-23. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3092300.

Mian, A., A. Sufi, and E. Verner. 2015. "Household Debt and Business Cycles Worldwide." NBER Working Paper 21581.

Okun, A. 1962. "Potential GNP: Its Measurement and Significance." In *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, 98–104.

Olley, G.S., and A. Pakes. 1996. “The Dynamics of Productivity in the Telecommunications Equipment Industry.” *Econometrica* 64 (6): 1263–97.

Reifschneider, D., W. Wascher, and D. Wilcox. 2015. “Aggregate Supply in the United States.” *IMF Economic Review* 63 (1): 71–109.

Reinhart, C., and K. Rogoff. 2009. *This Time Is Different: Eight Centuries of Financial Folly*. Princeton, NJ: Princeton University Press.

Rogers, J., C. Scotti, and J. Wright. 2014. “Evaluating Asset-Market Effects of Unconventional Monetary Policy: A Multi-Country Review.” *Economic Policy* 29 (80): 749–99.

Schivardi, F., E. Sette, and G. Tabellini. 2017. “Credit Misallocation during the European Financial Crisis.” BIS Working Paper 669. <https://www.bis.org/publ/work669.htm>.

Summers, L. 2014. “Reflections on the ‘New Secular Stagnation Hypothesis’.” In *Secular Stagnation: Facts, Causes, and Cures*, edited by C. Teulings and R. Baldwin. CEPR Press eBook. <https://voxeu.org/content/secular-stagnation-facts-causes-and-cures>.

World War II, the US Recovery from the Great Depression of the 1930s, and Postwar Productivity and Prosperity

Richard Sylla

Professor Emeritus of Economics and the former Henry Kaufman Professor of the History of Financial Institutions and Markets, New York University Stern School of Business

Introduction

The greatest economic crisis in the history of the United States was the Great Depression of the 1930s. The Depression proper is usually dated as the years 1929–1933, although in retrospect the entire decade 1929–1939 was one of approximately zero economic growth, and unemployment at the end of that decade was still in the double-digit range.

Following the Depression decade, the United States experienced the greatest economic recovery in its history. By 1943, unemployment was negligible and lower than it has been ever since that era. The reason, of course, was World War II, which led to a command economy with massive US government deficit spending and about 10 million of some 140 million Americans engaged in military rather than civilian economic pursuits.

The war began in Europe in 1939 and ended both there and in the Pacific in 1945. The United States did not formally become a belligerent until December 1941, after the Japanese attack on the American naval base at Pearl Harbor in the Hawaiian Islands followed by Germany's declaration of war on the United States. From 1939 to 1945, nominal US GDP grew at a rate of 16% per year, and real GDP at 11%. Per capita annual growth was 15% nominal and 10% real. The double-digit unemployment of the 1930s gave way to double-digit economic growth during the war years.

After the war and its massive fiscal stimulus ended in 1945, to the surprise of many economists the depressed economic conditions of the 1930s did not return. Instead, for decades after the war, the US economy experienced high levels of productivity growth and economic growth. Indeed, for the remainder of the 20th century, the United States experienced the highest rates of economic growth, as measured by real GDP per capita, in its history.

This chapter was produced for the 2018 World Economic History Congress session “Recovery from Large-Scale Crisis,” Boston/Cambridge, 2 August 2018.

What role did the war play in the recovery from the depressed 1930s? This is an old question with an old, usual answer. That answer says that the war provided a test of the theories developed by the great British economist John Maynard Keynes in the 1930s, especially Keynes's contention that a massive fiscal stimulus on the part of government could and would end a lingering economic depression. The prosperity of the United States during and after World War II was taken by Keynes's increasing number of disciples during the 1940s and 1950s as evidence that their master's theories passed the test with flying colors.

A second and newer question asks what role the war played in the decades of postwar prosperity. In other words, were there long-term effects of the war on prosperity as well as the shorter-term effect of recovery from the crisis of the 1930s? Robert J. Gordon in his 2016 book, *The Rise and Fall of American Growth*, answers in the affirmative.²⁹ He identifies a “Great Leap Forward” of the American level of labor productivity that occurred in the middle decades of the 20th century, which was “one of the greatest achievements in all of economic history” (p. 535). And Gordon also identifies what he deemed the source of the rise in productivity and economic growth: “World War II itself was perhaps the most important contributor to the Great Leap.”

It seems clear that many analysts regard World War II as an event of great importance in both bringing about the US recovery from the depressed economic conditions of the 1930s and ushering in decades of high economic growth after it occurred. But how were those achievements realized? This essay takes a fresh look at these issues.

A Somewhat Personal Note

As much of this history took place in my lifetime, let me begin with some personal experiences. I was born in 1940, some months after World War II began in Europe in September 1939. I have few memories of the war itself as it was taking place from 1940 to mid-1945. I do vaguely remember a third-birthday party (maybe from a film of it seen later), being quarantined with scarlet fever in 1944, my father going off to rifle practice on weekends, and the death of President Franklin Roosevelt in April 1945.

After the war, I remember our family purchasing some war surplus items, including binoculars that I still use. Sometime not long after the war, *Life* magazine published a volume called *Life's Pictorial History of World War II*, which I studied intently. Our family purchased our first television set in the

²⁹Robert J. Gordon, *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War* (Princeton, NJ: Princeton University Press, 2016).

late 1940s, and I shortly learned a lot more about the war from telecast documentaries such as *Crusade in Europe* and *Victory at Sea*. In the late 1940s there were lots of movies about the war, such as *The Fighting Seabees* and *The Best Years of Our Lives*, and I attended a number of them. I think this was when I first became interested in history.

During the first two postwar decades, 1945–1965, I transitioned from kindergarten to grad school. The United States was quite prosperous in these two decades, which I felt at the time and later had confirmed when as an economic historian I studied the era and compared it with the depressed 1930s and the great inflation of 1966 to 1981. But I didn't connect any of this with the war, which ceased to have the fascination it had when I was a boy. After 1945, the Cold War replaced World War II as a major geopolitical concern, and gradually from further study I became aware of how the two were related.

The Cold War ended (or seemed to) in 1989–1991, and I remember President George H. W. Bush noting that event and saying, “We won.” Economically, the 1990s were a prosperous decade, like most of those after 1940. In fact, as will be discussed further, the period 1940–2000 was the one that featured the highest economic growth of any six-decade period in US history. Since 2000, economic growth has slowed markedly, and the optimism of the 1990s has given way to quite a bit of pessimism about current economic conditions, the slow recovery from the financial crisis of 2007–2009, and the outlook for the US economy in the decades ahead.

Could the high-growth era of 1940–2000 have been rooted in World War II? That thought did not occur to me during the part of my academic career that took place in it, say, from 1963 to 2000. And at the end of the 20th century, no one was predicting that growth would slow. In fact, around the year 2000, the Clinton administration and Congress produced the first federal budget surpluses in many years.

Under the assumptions that the fiscal responsibility of the late 1990s would continue and that US economic growth would proceed at the high rates of the decades up to 2000, there were even rosy forecasts at the start of the 21st century that the country would be able to retire the entire national debt, then about \$5 trillion, by around now, 2018. Such a feat hadn't happened since Andrew Jackson's presidency in the 1830s.

But it was not to be. Economic growth slowed in the early 2000s. And a combination of tax cuts, war spending, and the only major US financial crisis of my lifetime caused the national debt, instead of being greatly reduced or eliminated, roughly to quadruple since 2000.

All this may seem quite remote from World War II more than seven decades ago. But another personal experience, a small one, led me to wonder

if in fact the war experience long ago is really so remote and irrelevant to the way we live now. Early in the 2000s, I began to notice that the ink on my morning hard copy of the *New York Times* seemed blurry, and since I was living in New York City, I assumed it must have been because the ink hadn't had time to dry before the paper was delivered to my door. After a few months, however, it dawned on me that it was not the fault of the *Times*, but of changes in my eyes. A visit to an ophthalmologist confirmed that I had cataracts, i.e., cloudy lenses. First, I had one original lens replaced with a plastic lens, and then the other a couple of years later.

During these procedures, the ophthalmologist explained to me the history of cataract surgery. It began in World War II, he said, when some US fighter pilots returned from their missions with shards of plastic in their eyeballs, the result of enemy fire which had shattered the clear plastic in the canopies of their aircraft. The doctors treating them soon noticed that there was no adverse reaction of the pilots' eyes to the plastic shards in them. Some of these doctors put two and two together and reasoned that if an eye could stand plastic in it, then plastic lenses could be inserted into eyes to replace the original-equipment lenses plagued with cataracts. That was the start of cataract surgery.

Six decades later, such surgery made it possible for me to keep on seeing clearly. And to keep on working for a decade after I turned 65, which I might not have been able to do with blurred vision. I suppose the discovery that lenses with cataracts could be replaced with plastic ones might have happened without World War II, but in fact it was the war that made it happen. The war experience may have speeded up the process of discovery.

My experience with cataract surgery, trivial as it was, made me wonder if there were other long-term effects of what happened during World War II, and if they might have contributed to the high-growth decades that followed. Was the development of other technologies of economic significance after 1945 speeded up by the war? That is one big question discussed below, in an admittedly preliminary way.

Another question (and the one that got me into this line of research) is how the war ended, not merely interrupted, the depressed economic conditions of the entire 1930s. I start with that recovery.

Recovery from the Depression

A simple story of the economic effects of World War II, often told by economists and historians in the decades after it, is that the massive wartime expenditures of the US government proved that John Maynard Keynes's macroeconomic theories and recommended policies were correct. Keynes's

General Theory of Employment, Interest, and Money, published in 1936 and inspired by the Great Depression of 1930–1933, argued against classical and neoclassical economic theories that asserted that a market economy would always right itself from a downturn and less-than-full employment via wage and price adjustments. Keynes demonstrated in theory that such a recovery might not happen and that a market economy could get stuck in a less-than-full employment equilibrium. That seemed to be the case during the 1930s.

To restore full employment, Keynes posited, a government would have to undertake a massive fiscal stimulus to increase aggregate demand to the point where it would match the economy's ability to supply goods and services. World War II in the simple story of recovery from the Great Depression provided the occasion to do just that, and perhaps a bit more than that.

The actual story of recovery is more complex than the simple story. There was a substantial recovery from the depths of the Great Depression during the 1930s, before the war began. It is told in the growth statistics. Real GDP per capita, the best measure of economic growth, fell from \$8,669 in 1929 (all data in 2009 dollars, from the website MeasuringWorth.com) to \$6,192 in 1933, a decline of 8.07% per year and an overall decline of 28.6%. Then it rebounded to \$8,643, little different from 1929, by 1937, a gain of 8.69% per year from 1933; and to \$8,881, slightly above 1929, by 1939. True, there had been little to no growth for a decade, and the unemployment rate was still in double-digit territory in 1939 (but far lower than in 1932–1933). Still, there was a substantial recovery.

Some perspective on the recovery of the 1930s can be gained by comparing it to the recent experience of the Great Recession of 2007–2009 and its aftermath. Real GDP per capita fell from \$49,300 (again, in 2009 dollars) to \$46,930 in 2009, a decline of 4.8% as compared with 28.6% from 1929 to 1933. It did not get back to the 2007 level until 2013 (\$49,317), six years later. Over US history since 1790, growth averaged 1.72% per year. From 1929 to 1939, it was anemic 0.24% per year, and from 2007 to 2016, a slightly less anemic rate of 0.53% per year. The Great Recession did not result in anything near the drop in GDP per capita of the Great Depression, and the recovery from it during the decade after it began, while anemic, was a bit more rapid than that of the 1930s.

The Great Depression of the 1930s should have ended in 1933, just as the Great Recession ended after a couple of years in 2009, but it didn't. A key difference between the Great Recession and the Great Depression was the policy response. Roosevelt's New Deal starting in 1933 did engage in deficit spending, but later research (by E. Cary Brown and others) indicated that there was little to no fiscal stimulus because if the economy had been at full

employment, the federal budget would have been balanced or even in surplus. Cary Brown said of fiscal policy in the 1930s that it didn't work because it wasn't tried.³⁰ The Obama administration's 2009 fiscal stimulus program of some \$700 billion, though poorly designed and later admitted to be inadequate, was better. But then the political stalemate in Washington after the 2010 elections prevented any further fiscal stimulus.

Innovative monetary policies, we now know, saved the day in the recent crisis. The Federal Reserve under Ben Bernanke engineered a fivefold expansion of the central bank's balance sheet by purchasing government and mortgage-backed securities and reduced its policy rate to nearly zero. That propped up asset prices, and the unemployment rate fell gradually from roughly 10% in 2009 to the 4% to 5% range (latest is 4.1%), close to full employment, now.

Monetary policy was also accommodative during the recovery from the Great Depression up to 1936, but its sources were different and the policy response was less enlightened than recently. Roosevelt's devaluation of the dollar in 1934 (a change in the gold value of the dollar from \$20.67 to \$35 per ounce) led to an inflow of gold and monetary expansion. So did the increasingly ominous situation in Europe, which caused Europeans to send more gold to the United States for safekeeping. Monetary expansion from these sources contributed a lot to the strong recovery from 1933 to 1936.

Then the Fed in 1936, unlike its more recent counterpart, became worried that the monetary ease it had fostered would lead to inflation unless something was done about it. (The Bernanke and then Yellen Fed, in contrast, worried that there wasn't enough inflation, so it maintained its low-interest and quantitative easing programs until recently.) The Banking Act of 1935, moreover, had given the Fed a new power to combat inflation by letting it raise the reserve percentage banks were required to hold. The Fed decided to try out its new tool by raising reserve requirements three times in late 1936 and 1937. Banks responded by cutting their lending, as they were supposed to do.

The result was a sharp recession in 1937–1938, in the context of an economy still with double-digit unemployment levels and far below its potential output. Monetary tightening may not have been the only culprit in this tale. Some contend the Wagner Act of 1935, which strengthened the bargaining power of labor unions, also contributed by the negative reaction it engendered in business leaders. And the Roosevelt administration, reacting to negative commentary on its fiscal deficits in the election year of 1936, decided to cut back on its fiscal stimulus programs, which as Brown and others argued were

³⁰E. Cary Brown, "Fiscal Policy in the 'Thirties: A Reappraisal," *American Economic Review* 46, no. 5 (December 1956): 857–79.

not really stimulating even before the cutbacks. Further, the introduction of Social Security in the late 1930s taxed workers and employers before any benefits were paid to retirees. That was not helpful, but the amounts seem too small to have much of a contractionary effect.

Together these policy shifts caused a serious recession in 1937–1938, interrupting what had been a strong recovery from the depths of 1932–1933. Hence the 1930s went down in history as “the Depression Decade,” with unemployment at double-digit levels throughout.

Nonetheless, the recovery resumed, and sharply, after the policy-induced recession of 1937–1938, in which real GDP per capita fell 4%. From 1938 to 1941, the last non-war year for the United States (the Japanese attack on Pearl Harbor coming at the end of the year), real GDP per capita grew at the very strong rate of 10.4% per year. Although the United States was not “at war” in these years, World War II, which officially began in September 1939, appears both in prospect and in retrospect to have caused this strong growth.

In prospect it did so because US leaders seem to have known for years that war was coming. During the mid- to late 1930s, there was the Spanish civil war; the Italian invasion of Ethiopia; German saber rattling, annexations, and appeasement at Munich; and Japan’s invasion of Manchuria. In a recent book, *Destructive Creation: American Business and the Winning of World War II*, Mark Wilson demonstrates that the US government and its military establishment were engaged in war planning and a major military buildup during these prewar years, and further, that American business was benefiting from strong European (mainly British and French) demand for war-related goods and services, and even from capital investments the Europeans provided to American firms to finance the facilities needed to produce what the Europeans wanted to purchase.³¹

After the war started in September 1939, and especially after the fall of France and the blitz of Britain in mid-1940, the United States accelerated its military buildup. Mark Wilson (2016, pp. 60–61) cites an estimate that “between June 1940 and December 1941, \$64 billion was spent and promised for defense production . . . a third of the total amount that would be spent by 1945.” He also notes that by mid-1941, half a year before the United States entered the war, the unemployment rate was down to about 4%. In a sense, the Depression of the 1930s was over before the United States entered the war. But the prospect of war was responsible for it.

The war itself replaced the double-digit unemployment rates of the 1930s with double-digit rates of economic growth. From 1939 to 1945, US nominal

³¹Mark R. Wilson, *Destructive Creation: American Business and the Winning of World War II* (Philadelphia: University of Pennsylvania Press, 2016).

GDP grew at a rate of 16% per year, and real GDP at 11%. Per capita rates were about 1% lower, but still double-digit. The war also brought a command economy from 1942 to 1945, when about a third of GDP consisted of production for the military. In the peak war years 1943–1945, 40% of the US labor force was devoted to military pursuits. The command economy led to production miracles that supplied both US and Allied forces. Robert Higgs succinctly describes them:

In 1940 and 1941 the economy was recovering smartly from the Depression, but in the latter year the recovery was becoming ambiguous, as substantial resources were diverted to war production. From 1942 to 1944 war production increased rapidly. Although there is no defensible way to place a value on the outpouring of munitions, the physical dimensions are awesome. From mid-1940 to mid-1945 munitions makers produced 86,338 tanks; 297,000 airplanes; 17,400,000 rifles, carbines, and side arms; 315,000 pieces of field artillery and mortars; 4,200,000 tons of artillery shells; 41,400,000 rounds of small arms ammunition; 64,500 landing vessels; 6,500 other navy ships; 5,000 cargo ships and transports; and vast amounts of other munitions. Despite administrative mistakes, frustrations, and turf battles, the command economy worked. But, as always, a command economy can be said to work only in the sense that it turns out what the authorities demand. The U.S. economy did so in quantities sufficient to overwhelm enemy forces.³²

Postwar Prosperity

When the war ended in 1945, the winding-down of its fiscal-military stimulus, which had already begun in 1944, led some knowledgeable observers to predict that the depressed economic conditions and high unemployment of the 1930s were likely to return. It didn't happen. There were a number of reasons.

A long-popular one was “pent-up domestic demand.” American workers were fully employed, perhaps even more than fully employed with extra shifts and overtime during the war, but there wasn't much for them to buy. My father bought a new 1941 Chevrolet just before the United States entered the war. From 1942 to 1946, there were no such cars produced. In 1947, Chevrolet began to produce and market cars again. The 1947 model looked a lot like the 1941 model that served as our family car until 1950. The main difference was that the 1947 car sold for twice as much as the 1941 model, for which my father had paid \$900. The US price level roughly doubled between 1939 and

³²Robert Higgs, “Wartime Prosperity? A Reassessment of the U.S. Economy in the 1940s,” *Journal of Economic History* 52, no. 1 (March 1992): 41–60, quote at 55–56.

1948, with most of the inflation coming after the war as most price controls were removed.

So, despite higher rates of taxation during the war and the innovation of tax withholding from paychecks, Americans managed to save a lot of money, some 20% to 21% of after-tax disposable income deposited in banks and invested in patriotic government bonds. These funds became available to spend on consumer goods after the war, once demobilization was completed and production resumed for non-war goods. A high level of consumer demand was there to prevent a return to Depression conditions.

Another, more recent explanation is foreign demand. Net exports (exports minus imports) were unusually high as a percentage of GDP in the immediate postwar years, 1946 and 1947, in part because the production facilities of the United States were largely intact, unlike the facilities of either US enemies or allies during the war. About half the jobs created in 1946 and 1947 can be attributed to the employment effects of the rise of net exports. These demands could be and often were financed by loans from American lenders.³³

Then, after conversion from wartime to peacetime production, US net exports continued, albeit at less lofty percentages of GDP, in subsequent years with Marshall Plan and other financing from the United States. Often US lending and Marshall Plan financing were tied to purchasing from US suppliers, which helped to ease the transition from war-related to civilian spending. In reality, foreigners had few options other than purchasing from the United States because so many other economies were devastated and disrupted by the war. Net exports from the United States remained positive for more than two decades after the war, but they have been mostly negative ever since.

Banks and other financial institutions in the United States ended the war with huge amounts of government debt on their balance sheets. They were in a strong position to provide financing to Americans for the houses, cars, and appliances that Americans demanded after the war. All they had to do to gain the wherewithal for such lending was to liquidate their bond holdings, which they did for many years. Bankers described it as getting back into the business of banking, as opposed to financing the government's war efforts. An expanding supply of credit, very unlike the conditions of the Great Depression, contributed to postwar economic expansion.

Another reason Depression levels of unemployment did not return after 1945 had to do with changes in the labor force. During the war, when 10 million potential civilian workers were in the military forces, large

³³Jason E. Taylor, Bharati Basu, and Steven McLean, "Net Exports and the Avoidance of High Unemployment during Reconversion, 1945–1947," *Journal of Economic History* 71, no. 2 (June 2011): 444–54.

numbers of females entered the labor force, as exemplified by the famous “Rosie the Riveter.” After the war, when the military downsized its forces, which possibly might have created civilian unemployment, most of the Rosies left the labor force, many becoming housewives and mothers to the postwar baby boomers.

Congress helped, too, by enacting what was called the GI Bill, which sent many veterans to college instead of returning immediately to the labor force. GI benefits also extended to Veterans Administration mortgage loans to help returning GIs buy houses, which helped to sustain a high demand for housing. When I bought a house in 1970, I assumed a 6% VA mortgage loan, then a below-market rate. So even though I never served in the military, I received a small benefit from a war-related program.

Finally, the Cold War, which appeared shortly after World War II ended, provided a reason to sustain government spending on military preparedness at a high level for decades. The United States created a “military-industrial complex,” which President Eisenhower so named and expressed some misgivings about at the end of his presidential term in 1961. Whatever their dangers, the Cold War and the military-industrial complex provided continuing fiscal stimulus to the American economy for decades. During those decades, recessions were mild, few, and far between. Unlike after earlier wars in US history, after World War II the national debt increased instead of falling. The rate of increase was modest in the first postwar decades while the economy expanded at high rates. The debt/GDP ratio therefore declined into the 1970s.

Rapid Economic Growth after World War II: A Result of the War?

My Harvard College classmate and fellow economist Robert J. Gordon, as noted earlier, published an excellent and provocative 2016 book, *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War*. The book received much media attention for its pessimistic conclusion that America’s best economic days were behind it and that growth, already slowing for a decade or more when the book appeared, would continue at considerably lower than past historical rates for the foreseeable future.

According to Gordon, the United States had a “glorious century” from 1870 to 1970 propelled by great innovations such as public water supply and indoor plumbing, internal combustion engines and motor vehicles, electricity and electronic innovations, mass education, and so on. The great innovations increased labor productivity and total factor productivity (TFP, defined below) to higher levels, spurring economic growth. The great innovations of the

glorious century were much more important in raising productivity than the more recent innovations that so fascinate us, such as information and telecommunication technologies, the internet, Google, Facebook, and smartphones. To clinch his point, Gordon would ask audiences to consider if they had to choose between having a smartphone and having indoor plumbing. Which would they choose? Overwhelmingly the answer was indoor plumbing.

In addition to the glorious century, 1870–1970, Gordon identified the “Great Leap Forward” that I mentioned earlier, which, as noted at the outset of this essay, he attributed to World War II and deemed “one of the greatest achievements in all of economic history.”

It is time to define some terms. Economists break down output per capita (and by extension, its growth rate) into its proximate determinants. One version of this, designating total output (GDP) as Q , population as N , hours worked as H , and the labor force as L , is that output per capita is

$$Q/N = Q/H \times H/L \times L/N.$$

Or, in words, output per capita consists of output per hour worked (labor productivity) times hours per worker (the work year) times the labor force participation rate (the percentage of the population that is in the paid labor force). Growth is then measured by how much Q/N and its proximate determinants change over time.

Since hours worked and the labor force participation rate tend to change slowly over time, in this formulation it is changes in labor productivity (Q/H) that are usually the main determinant of growth. In fact, H/L (the work year) usually declines over time as people take part of their rising incomes in the form of greater leisure. Arithmetically, such declines subtract from growth, but some economists argue that a reduction in hours, H , actually increases labor productivity (because, for example, workers are more rested when they work fewer hours). So there can be interactions among the proximate determinants of growth.

Another version of the determinants of growth defines total output as the consequence of total factor productivity (TFP), which is real output divided by weighted total inputs of (usually) labor (L) and capital (K), with weights of about 0.7 for L and about 0.3 for K , based on the approximate long-term shares of L and K in Q . Divide total output by population to get real per person GDP, the growth of which over time is generally considered the best measure of economic growth. In this version of growth analysis, TFP is the major determinant of long-term growth, with capital deepening, that is, more capital per worker and per member of the population, also contributing.

Chapter 16 of Gordon's book, entitled "The Great Leap Forward from the 1920s to the 1950s: What Set of Miracles Created It?" documents dramatic increases in both labor productivity and total factor productivity that occurred around the time of World War II. Compared to a long-term trend growth of 1.9% per year during the period 1870–1928, labor productivity—output per hour of work—was on the trend line in 1935. Then in the buildup to the war it increased to 11% above trend by 1941. It "then reached 32 percent above trend by 1957 and 44 percent above trend in 1972. The post-1928 growth miracle is perhaps the central puzzle in the American economic history of the 20th century" (Gordon 2016, p. 539). After 1972, labor productivity stays above the trend line established for 1870–1928 but is roughly parallel to it. Therefore, there was nothing unusual about labor productivity growth after the early 1970s. If anything, Gordon's Figure 16-1 appears to indicate that productivity growth was a bit less after the early 1970s than it was from 1870 to 1928.

Gordon's TFP data, indicating the growth of output per weighted unit of labor and capital input, draw even more attention to the exceptional nature of the World War II era. The data are presented by decades, with the four highest decades of TFP growth from 1890 to 2014 being, in order from top to bottom, the 1940s (3.4%), the 1930s (1.8%), the 1950s (1.6%), and the 1960s (1.4%). With the exception of the 1920s (1.3%), in all the other decades from 1890 to 2014, TFP growth was less than 1% per year.

Gordon's analysis of why the Great Leap Forward in productivity occurred around the time of World War II gives some credit to the Great Depression and the New Deal. Competitive pressures to cut costs forced businesses to pursue lean production using less labor to produce output in the Depression decade. Then the New Deal added to the nation's infrastructure with its spending on dams, bridges, and highways. And New Deal labor legislation strengthened labor's bargaining power, raising wages and reducing hours. Even earlier, in the 1920s, restrictions on immigration and high tariffs had begun to increase labor's clout. These developments of the 1920s and 1930s encouraged business to substitute capital for labor, raising labor productivity after the war.

For Gordon, the war itself was even more important. The government spent huge amounts of money on plant and equipment, much of it operated by private business, and encouraged high rates of production by buying the output of the new production facilities. High production rates lowered unit costs via learning-by-doing, notably in the aircraft and shipbuilding industries. Once learned, Gordon says, the lessons of wartime production were not forgotten. Labor productivity and TFP growth rates, having leapt well above their earlier trend, remained there for several decades.

Much of Gordon's story, however, is backward-looking rather than forward-looking. For him, the great innovations occurred earlier, as far back as the late 19th century, during what some call the Second Industrial Revolution when electricity and internal combustion came in as general-purpose technologies. What the Great Depression and World War II did was to raise the efficiency of producing all the varied offshoots of those general-purpose technologies to a much higher level, leading to higher rates of economic growth.

Gordon does not ignore new technologies that may have had roots in World War II developments. He does discuss Third Industrial Revolution technologies such as computers, communications, other digital technologies, and advances in medicine (recall my cataract surgery). His argument is that while there has been a rapid pace of innovation in these areas, they have far less impact on the growth of labor productivity and TFP than did the great innovations of the Second Industrial Revolution.

The slowing of productivity growth since the 1970s is Gordon's best evidence of the lesser impact of new technologies, although it did not seem to slow economic growth much if at all during the last three decades of the 20th century. To show that I studied the growth record of the United States since 1790, using the data and calculators at MeasuringWorth.com, and asking what were the highest periods of economic growth in U.S. history.

Since 1790 (1790–2016, 2016 being the last year available), economic growth in the United States has averaged 1.72% per year. The slowing of growth since 2000 makes little difference in the long-run rate; from 1790 to 2000, the rate was 1.77%. So, we can feel comfortable in saying that the long-run rate has been about 1.7% per year.

With the long-run of 1.7% as a benchmark, we can ask, What were the periods (of various lengths) of highest economic growth in the 23 decades of US economic history after 1790? Here is the answer:

	Highest Growth (annual rate)	Second Highest (annual rate)
7 decades	1930–2000, 2.51%	1920–1990, 2.36%
6	1940–2000, 2.59%	1930–1990, 2.56%
5	1940–1990, 2.67%	1930–1980, 2.60%
4	1940–1980, 2.75%	1930–1970, 2.73%
3	1940–1970, 2.96%	1930–1960, 2.65%
2	1930–1950, 3.08%	1940–1960, 2.97%

It is interesting, I think, that all these decades include the World War II period and that (with one exception, 1930–1960, which had slightly lower annual growth than 1930–1970) the shorter the period that includes World War II, the higher was the rate of growth. The highest rates of growth came in the 2 two-decade periods that include the war. And as one increases the number of decades from two to seven, the growth rate declines almost monotonically, but it remains at levels well above the US long-term average of about 1.7%. The World War II era does indeed seem special.

Using the MeasuringWorth.com data, Gordon's glorious century of economic growth, 1870–1970, featured a growth of real income per capita of 2.05% per year. From 1970 to 2000, the rate was 2.22% per year. Growth did not slow after Gordon's glorious century; it increased a bit. Since productivity growth slowed and hours did not increase, continued high growth from 1970 to the end of the 20th century must have been attributable to gains in labor force participation as baby boomers, male and especially female, entered the labor force. That seems plausible, and it also points to reasons why growth has slowed in the 21st century: the baby boomers who swelled the American labor force during the 1970s and 1980s have started to retire in recent years, and the large gains in female labor force participation are a thing of the past.

Gordon's explanation of the great leap forward of productivity in the World War II era of the middle decades of the 20th century is a bit vague. But it seems to boil down to a view that the great general-purpose technologies of the Second Industrial Revolution that began in the late 19th century, notably internal combustion and electrification, reached the height of their development aided by wartime demands. Newer technologies developed during and after the war right down to the present, in Gordon's view, were simply less capable of increasing labor and total factor productivity, however else they may have enriched our lives.

An alternative, more historical explanation is, to my mind, equally plausible based on productivity growth patterns, which were high and rising from the late 1930s, when war demands kicked in, to the early 1970s, when productivity growth reverted to prewar levels. It is that the war catapulted the United States to a unique position in modern economic history. At the war's end, the productive capabilities of the United States were greatly enlarged and undamaged. Virtually every other major economy, whether an enemy of the United States or its ally, was devastated. Hence, in 1945 the US economy produced about half of the world's total output and total industrial production.

Unique circumstances produce unique outcomes. The rest of the world, much of it devastated, needed what the United States could and would produce as it rebuilt. Current account surpluses were persistent from the war to

the 1970s, and they were used to finance the recovery of Europe and Japan by early Cold War military spending around the world, by US corporate spending and American bank lending overseas, and by outright grants such as the Marshall Plan, often with requirements that the grants be spent on American products. All this generated considerable foreign demand for the output of American industry and agriculture.

At home, demand was also high. Americans came out of the wartime command economy with considerable savings, and American financial institutions were flush with government bonds they would gradually run down to finance bank loans to the private sector. Labor unions were in a strong bargaining position with American firms and industries that faced limited foreign competition for two to three decades. Therefore, postwar prosperity was widely shared, in contrast to recent years when the proceeds of growth rebound mostly to the benefit of the top 1% or 10% of income earners.

These conditions were gone by the 1970s, if not before. Europe and Japan had fully recovered after a similar period of high economic growth, and their corporations and farmers were in a position to compete with American producers. The worldwide dollar shortage of the 1940s and 1950s became a dollar surplus in the 1960s, and by the early 1970s it toppled the Bretton Woods international monetary system created during World War II. Americans saved less and consumed more. The current account surpluses of the quarter-century after the war shifted to deficits in the 1970s. The bargaining power of American labor declined as manufacturing shifted from the Rustbelt to the Sunbelt and overseas, while also becoming a relatively smaller part of a post-industrial economy in which services steadily increased in their share of total output.

For all these historical reasons, should we be surprised or puzzled that the high levels of US labor productivity and TFP experienced from the late 1930s to the late 1960s or early 1970s reverted to more normal rates of growth? I don't think so. The war created them, and the results of the war sustained them for roughly another quarter-century. Then unique circumstances reverted to more normal circumstances.

Conclusion

Anticipations of World War II in the late 1930s promoted recovery from a decade of depression before the war began, but only because ill-timed US contractive economic policies in the mid-1930s led to a major recession in 1937–1938, stifling a strong recovery from the Great Depression's 1933 depths. Those anticipations led European nations and American military planners to place orders for military supplies from US producers. As a result,

the US economy reached substantially full employment before the United States entered the war at the end of 1941.

The command economy that followed from 1942 to 1945 achieved remarkable productive feats in supplying US and Allied forces with the weaponry, food, clothing, and shelter they needed to defeat the Axis powers. High taxes and rationing restrained demand for non-military goods and services. Government borrowing swelled the national debt to 120% of GDP, still an all-time record for the United States, while also raising the assets of US households and financial institutions. Productivity soared to high levels that were maintained in civilian production long after the war ended in 1945.

After 1945, most of the forces that sustained demand and prevented a return to the depressed condition of the 1930s lasted for at least a couple of decades. A positive level of net exports was one such force. During most of that era the rest of the world needed what the United States, and no other nation, could supply, including both goods and the credit needed to buy them. American families were generally prosperous, with less inequality of income distribution than characterized the period before the war or since approximately 1980, and they used that prosperity to sustain high levels of spending on consumer goods and to have a lot of children, the so-called baby boomers. Credit flowed. American business corporations, with little competition from foreign firms, earned good profits, and the stock market recovered. The benchmark Dow Jones Industrial Average after a quarter-century regained 1929 levels in 1954, and then went on to more than double 1954 levels in the early 1960s. And women accepted a lower rate of labor force participation than they had during the war, which meant there were plenty of jobs for returning veterans, whether they went right to work or delayed that for a time while in college on the GI Bill.

There remain many questions about how the high levels of productivity during World War II were achieved, and then sustained for decades. But the burden of proof is on those who would deny a connection of the US war experience to higher productivity, and that burden is substantial.

The Age of Experts: A Review of Marc Levinson's *An Extraordinary Time*

Stephen Sexauer

Chief Investment Officer, SDCERA

Laurence Siegel

Gary P. Brinson Director of Research, CFA Institute Research Foundation

How good were the good old days? Was the post-World War II economic boom in the United States and other developed countries a truly special period, one that we cannot expect to repeat, even over centuries-long time frames? Where did those exceptional growth rates come from, and what—if anything—can we do to bring them back?

In *An Extraordinary Time: The End of the Postwar Boom and the Return of the Ordinary Economy*, the economic historian and journalist Marc Levinson, author of *The Box*, poses these questions. His answer is that the quarter-century from 1948 to 1973 was truly exceptional and that the good times are not coming back. They were, he writes, “an economic golden age across the world” [inside cover] and, because of the rapid rise in oil prices in the 1970s and the economic ups and downs since then, “we are not likely to see its equivalent again” [p. 270].³⁴

Levinson presents a richly detailed account of the development of macroeconomic models and their use in the postwar period by governments trying to harness the chaos of economic innovation and growth. That tale is the sterling contribution of this book: the emergence of the statutory role of government to pursue and attain economic growth and stability. The arrival of this era of *dirigisme* is now largely forgotten and Levinson recounts it effectively. But Levinson’s pessimism about future growth asks that we ignore essential elements of the human story and of economic history. The global economy made great progress before 1948 and after 1973; if you do not care only about people in the United States, the most recent quarter-century is the most special one of all. Levinson’s conclusion—that we are destined to revert to an

Reprinted from the *Journal of the National Association for Business Economics*, 26 October 2017, with the permission of the publisher, the National Association for Business Economics.

³⁴This latter quote is from the last sentence of the book.

“ordinary economy,” one characterized by stagnation and unproductivity—is deeply wrong.

Innovation versus Stability

As a general rule of economic progress: no change, no growth. The tensions and balances between governments, the entrepreneurial and professional classes, and rank-and-file workers ebb and flow depending on which economic condition dominates: innovation or stability. It is difficult to have both.

But, in the United States in 1948–1973, we found instead a period of innovation and income stability. The United States had earlier experienced rapid innovation, or growth, in many of the years between the Civil War and World War I, and then again in the 1920s. But the growth was usually chaotic and disruptive, with many people hurt by the side effects of innovation. The 1948–73 period was indeed special, in that the fast-growing US economy is remembered as relatively stable, with fewer people hurt, and the perception that social cohesion provided the lubricant for great change without great angst.

You’d be forgiven if you just scratched your head: *what* social cohesion? The civil rights battle raged over the entire period. In 1960, the sunny midsummer of our history, John Kennedy ran for president on a platform of “get[ting] this country moving again” following the 1958 recession, which was one of three over the previous eight years and was the deepest since the Great Depression. The chaos that reigned from the Kennedy assassination through the end of the Vietnam War more than a decade later is nobody’s idea of social cohesion. And that last period took up two-fifths of the special quarter-century; it was no brief interruption.

And that was just the United States. For much of the world, 1948–1973 was a turbulent and sometimes violent period. From the Soviet Union to the coups and kleptocracies in emerging markets, life was chaotic and desperately poor. As Levinson recounts: “There is no sugarcoating the brutality that, for many people . . . in many corners of the world, . . . was part of everyday life” [p. 44].

Levels Matter

Rates of change are felt more keenly by the human mind than levels. And some of the social cohesion in the 1948–73 period was real: As many as 80% of Americans saw themselves as middle class. They went to the same schools, spoke the same language, and shared the recent memory of a war that was almost lost.³⁵ Despite increasing urbanization, small-town living—where

³⁵We should not forget that most African Americans did not go to the same schools as the majority. Even after schools became legally integrated, most remained segregated due to living patterns.

people from all walks of life rub shoulders and are forced to get along—was still the predominant memory, if not the present reality.

But, setting aside emotion and invoking rationality, would you rather live in a \$15,000 GDP per capita country that is growing at 8% or a \$52,000 GDP per capita country that is growing at 2%?³⁶ We'd choose the latter, simply because we don't want to struggle. We can find emotional satisfaction in ways other than anticipating prosperity that may or may not come. Levels matter.

No Economic Period Is Ordinary

In the long sweep of human achievement across all peoples before 1948 and after 1973, the competing epochs of progress are also quite extraordinary. Levinson mostly pooh-poohs them, writing:

The effects of innovation on the economy were slight in the early twentieth century, very strong from the 1920s to 1973, quite weak between 1973 and 1995, fairly strong between 1995 and 2003, and considerably weaker in the years thereafter. . . . [I]n the late 1990s . . . the commercialization of the Internet [led] to a burst of productivity growth . . . that exhausted itself in just half a dozen years. [p. 263]

This observation brings us to our most basic objection to Levinson's theme: No economic period is ordinary. Was any time more special than the end of the 18th and the beginning of the 19th century, when substitutes for human muscle power, stored in carbon, were harnessed to begin the release of all humankind from a Malthusian existence? Was any time more extraordinary than the opening of China and the end of the Cold War, enabling two or three billion people to apply their skills in a world economy that had previously been the province of the Europeans, Americans, and Japanese?

What the Nobel Prize-winning economist Angus Deaton calls “the great escape,” from dangerous and backbreaking work in fields, mines, and primitive factories, came largely before 1948 in the industrialized West and is still proceeding today in emerging markets.³⁷ The drive to bring sanitation and clean water to the masses began early in modern history but continues to be a challenge in developing countries today.

Medical innovation has produced a decrease in suffering and gains in the quality and length of life that can hardly be comprehended; in 1840 surgery

³⁶US per capita GDP in 2017 dollars ranged between \$15,000 and \$16,000 in 1951–1954 and is \$52,000 today. Source: Federal Reserve Bank of Saint Louis, <https://fred.stlouisfed.org/series/A939RX0Q048SBEA>.

³⁷Angus Deaton, *The Great Escape* (Princeton, NJ: Princeton University Press, 2013).

had to be performed without anesthetics. (Let that sink in.) The first cure of an infectious disease using antibiotics was in 1935—and we have barely seen the beginning of these advances, which continue today with gene sequencing. Gains in human rights and political freedom, along with occasional losses, have also proceeded without reference to any particular quarter-century.

In the Postwar World, a New Leaf: Government by Experts

In making his case, Levinson does provide us with a valuable and richly detailed view into two powerful postwar trends. The first consists of the emerging availability of government data on the economy and the rise of operations research as a discipline that could be used by government experts for macroeconomic planning. The second is that governments made the leap to being accountable for economic outcomes, combining the need to please voters with the capacities newly acquired through macroeconomic data and theory.

This trend toward management of non-centrally planned economies by government experts is apparent from the wording of the US Employment Act of 1946, which “promote[s] maximum employment, production, and purchasing power.”³⁸ It lives on in the July 2012 statement by European Central Bank president Mario Draghi that the bank would do “whatever it takes to preserve the euro.”³⁹

Experts and the Magic Square

Levinson recounts the effort by experts to manage four macro variables to desirable and stable levels to achieve growth and stability, a strategy called the “magic square.” The variables were unemployment, inflation, real GDP growth, and trade balances.⁴⁰

In the United States, the reigning expert was Walter W. Heller, a Kennedy confidant and strong Keynesian.⁴¹ In the developing world, it was Raúl Prebisch, an Argentine economist who tried to steer a middle course between capitalist orthodoxy and the radicalism then in vogue in Latin America and other emerging economies.

³⁸http://www.federalreservehistory.org/essays/employment_act_of_1946.

³⁹https://www.nytimes.com/2016/01/22/business/international/ecb-sti_mulus-mario-draghi.html.

⁴⁰In addition, they were committed to maintaining fixed exchange rates, a system that failed by 1971; it’s a matter of interpretation whether the foreign exchange value of one’s home currency is a fifth variable or just a manifestation of trade balances.

⁴¹Not Walter E. Heller, the Chicago financier and philanthropist.

And in West Germany the rock-star expert was Karl Schiller, *Stern*'s Man of the Year in 1969, described by Levinson as “work[ing] late into the night [with] his ‘team of eggheads.’ After crunching the numbers, they specified the most desirable rate of economic growth,” which turned out to be 4%, with 0.8% unemployment and 1% inflation. The cadre of experts used linear programming, input–output analysis, and other new items in the economist’s toolkit to balance the unpredictable behavior of markets and capitalists against the legal obligation to foster growth.

In the end, however, such incredibly rosy macro goals were unachievable. As Levinson indicates, 1973, when the first large OPEC oil price increase occurred, was the year when the macroeconomic indicators all seemed to agree that things had started to go badly. Yet the failure in 1971 of the Bretton Woods fixed exchange-rate system proved that the magic square had been deteriorating for some time; OPEC was merely the catalyst.

By the time of the second oil shock, in 1979, any pretense of maintaining the magic square had been abandoned.

US inflation was running at 9%, unemployment was around 6%, and real GDP growth was a paltry 1.3%. President Carter may not have uttered the word “malaise” in a speech but he should have. In the end, Levinson states, “the idea that government planning could assure prosperity and rising living standards for all proved to be a cruel hoax” [p. 46].

We don’t think any economist today would propose that you can hold all of the magic square variables constant at the same time, when each is affected by millions of independent decisions and several of the variables are pushing against each other. (For example, some factors that cause high GDP growth tend to also cause high inflation.) “Only belatedly,” Levinson writes, “would [Karl Schiller] accept that the magic square was a technocrat’s fantasy” [p. 34]. Governments have learned their modesty lesson, although it could also be argued that the outcomes are also modest: In what seems like the main challenge we now face, First World economies are hardly growing faster than their populations.

The problem is not that the experts don’t know anything; they do. With a few exceptions, they are diligent and well-meaning students of the economy and of human behavior. The difficulty is that the rational expectations hypothesis is roughly right: People do what they want and react poorly to policies that push them to do things they think are unprofitable. As a result, there are severe limits to what proactive economic policy can accomplish. Mostly, we need to establish good laws and institutions and hope for the best, figuring that people acting in their own interest will produce the “right” amount of economic growth.

The Return of the Ordinary Economy?

Levinson's insights into the role of economic planning, in both developed and emerging economies, make *An Extraordinary Time* worth reading even if its conclusions are suspect. We'll close by noting that the last half of the book's subtitle, *The Return of the Ordinary Economy*, posits that there is such a thing as an ordinary economy. What is it, and what will it look like when it returns?

Perhaps, for many people, it is a steady-state economy where one day, year, or decade is much like another, lacking in abrupt surprises and limited in the amount of change. Unfortunately, the only steady-state economy in the historical record is one that, according to Angus Maddison, produced output equal to \$3 per person per day—just enough that many people could survive long enough to reproduce; a lucky few survived longer, and the tiniest minority did so at some degree of affluence. This was the reality faced by most of the world's people for most of its history.

This Hobbesian way of life, “solitary, poor, nasty, brutish, and short,” is, of course, not what Levinson thinks is in store for us. His ordinary economy looks more like today's economy: affluent yet sluggish, unequal, unfamiliar, and continually disruptive. Like other forecasters of our economic future, Levinson has projected the present forward indefinitely, making the familiar behavioral mistake of overemphasizing the most recent observation.

We have more faith in humanity than that. The evidence is on our side.

Measurement and Mismeasurement

Underestimating the Real Growth of GDP, Personal Income, and Productivity

Martin Feldstein

George F. Baker Professor of Economics, Harvard University

Economists have long recognized that changes in the quality of existing goods and services, along with the introduction of new goods and services, can raise grave difficulties in measuring changes in the real output of the economy. Prominent economists have led and served on government commissions to analyze and report on the subject, including the Stigler Commission in 1961, the Boskin Commission in 1996, discussed in a symposium in the Winter 1998 issue of this journal, and the Schultze Commission in 2002, discussed in a symposium in the winter 2003 issue of this journal (Stigler 1961; Boskin et al. 1996; National Research Council 2002). But despite the attention to this subject in the professional literature, there remains insufficient understanding of just how imperfect the existing official estimates actually are.

After studying the methods used by the US government statistical agencies as well as the extensive previous academic literature on this subject, I have concluded that despite the various improvements to statistical methods that have been made through the years, the official data underestimate the changes of real output and productivity. The measurement problem has become increasingly difficult with the rising share of services that has grown from about 50% of private-sector GDP in 1950 to about 70% of private GDP now. The official measures provide at best a lower bound on the true real growth rate with no indication of the size of the underestimation. Thus, Coyle (2014, p. 125) concludes her useful history of GDP by saying, “Gross domestic product is a measure of the economy best suited to an earlier era.”

In considering these issues, I have been struck by the difference between the official statistics about economic growth and how people judge whether their own economic condition has improved. The official figures tell us that real GDP per capita grew at an average rate of just 1.4% during the past

Reprinted from the *Journal of Economic Perspectives*, Spring 2017, with the permission of the publisher, the American Economic Association.

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.31.2.145doi=10.1257/jep.31.2.145>.

20 years. It is common to read in the press that because of this slow overall growth and changes in the distribution of income, the real income of the median household did not rise at all between 1995 and 2013 (for example, in the Council of Economic Advisers' *2015 Economic Report of the President*, p. 30). When polls ask how the economy is doing, a majority of respondents say the country is doing badly; for example, 57% of respondents to a CNN-ORC poll in January 2016 said that the country is "doing poorly" (as reported in Long 2016) and in a Gallup poll in October 2016, 29% of respondents said the US economy is "poor" while only 29% said it was good or excellent (as reported in Dugan 2016). But in a Federal Reserve (2014) survey of household attitudes, two-thirds of households reported that they were doing as well or better than they had been five years earlier and that they were either "living comfortably" or "doing OK."

The contrast is revealing. People know their personal experience directly, but they depend on news media, politicians, and official statistics to judge how the economy as a whole is doing. And while the government statisticians are careful to say that GDP doesn't measure how well we are doing, there is a strong temptation on the part of the press, the politicians, and the public to think that it measures changes in the real standard of living. In this way, when the official statistics on economic growth understate real economic growth, it reduces public faith in the political and economic system. For example, the low measured growth of incomes probably exacerbates concerns about mobility, with people worrying that they and their children are "stuck" at low income levels: in a CNN/ORC poll, 56% of respondents said they think their children will be worse off than they are (as reported in Long 2016), and in a Pew Research Center poll, 60% of Americans said their children will be financially worse off than their parents (at <http://www.pewglobal.org/database/indicator/74/survey/all/response/Worse+off/>). Moreover, I think it creates a pessimism that contributes to political attitudes that are against free trade and critical of our market economy more generally.

I begin this essay by briefly reviewing the age-old question of why national income should not be considered a measure of well-being. I then turn to a description of what the government statisticians actually do in their attempt to measure improvements in the quality of goods and services. Next, I consider the problem of new products and the various attempts by economists to take new products into account in measuring overall price and output changes.

Although the officially measured rates of output growth have slowed substantially in recent years, the problem of understating real economic growth

is not a new one.⁴² It reflects the enormous difficulty of dealing with quality change and the even greater difficulty of measuring the value created by the introduction of new goods and services. This paper is not about the recent productivity slowdown, but I return to that issue later in this paper and discuss the implications of these measurement issues for the measurement of productivity and the recent slowdown in the rate of productivity growth.

The final section of this paper discusses how the mismeasurement of real output and of prices might be taken into account in considering various questions of economic policy. Fortunately, there are important uses of nominal GDP that do not require conversion to real GDP.

Not Even Measuring Output, and Certainly Not Well-Being

There is a long-running debate about the extent to which national income estimates should be designed to measure the well-being of the population or just the output of the economy. But in practice, national income concepts have been intentionally defined in ways that fall far short of measuring even *economic* well-being, let alone the broader well-being of individuals as influenced by matters like the environment and crime.

Even if we focus just on economic output, the concept of national output has been explicitly defined ever since the initial work of Kuznets (1934) and Kuznets, Epstein, and Jenks (1941) to exclude goods and services produced within the home. An earlier National Bureau of Economic Research study by Mitchell, King, and Macaulay (1921) offered a conjectural value of housewives' services equal to about 30% of their estimate of the more narrowly defined traditional national income. Frazis and Stewart (2011) estimate that household production, under various assumptions, ranges from 31% to 47% of money earnings. The official statistics also exclude services that are provided outside the home but not sold. This omission has probably had a larger effect in recent years with the provision of such services as Google and Facebook and the vast expansion of publicly available videos and music, together with written commentary, stories, reports, and information, all of which are now available to web-connected users for essentially zero marginal payment.

Similarly, national income estimates focus on the positive value of the goods and services that households consume, not on the time and effort involved in

⁴²The vast literature bearing on the measurement of changes in the real output of the economy reaches back to Sidgwick (1883), Marshall (1887), Kuznets (1934), and Kuznets, Epstein, and Jenks (1941) and includes, more recently among others, Griliches (1992), Nordhaus (1997), Hausman (1996, 1999), and Gordon (2016). The NBER Conference on Research in Income and Wealth has focused work on this issue for more than 80 years (as discussed in Hulten 2015).

earning the funds to buy those goods and services. The average workweek has declined but the number of two-earner households has increased. Working conditions have improved as employment has moved from factories and farms to offices. All of this affects economic well-being, but there is (by agreement) no attempt to take it into account in our measures of national income.

I mention these issues not to criticize the official definition of national income, but to stress that it is intended by design to be a measure of national output, not a measure of well-being. The public clearly wants a description of changes in well-being and inappropriately uses the official measures of real GDP and real personal incomes for that purpose. It might be useful to develop a formal array of well-being indicators and perhaps some summary index. These indicators might include measures of health, air pollution in cities, crime, and other matters that are not measured in the official economic statistics: Coyle (2014, chap. 5) discusses some previous attempts to provide such additional indicators. Alternatively, more attention might be focused on the Federal Reserve's *Survey of the Economic Well-Being of U.S. Households* and its frequency might be increased from an annual survey to quarterly to increase its public saliency.

However, in this essay I will set aside the issues concerning what economic and noneconomic factors are left out of GDP, and how a broader measure of well-being might be constructed. Instead, I will argue that the official measure of real GDP does not even achieve its stated goal of measuring real national output on its own terms.

Measuring Quality Change

The government's calculation of real GDP growth begins with the estimation of nominal GDP, which is the market value of the millions of goods and services sold in the market to households, firms, governments, and foreign buyers. The government statisticians do a remarkable and prodigious job of collecting and then updating data from a wide array of sources.⁴³

But for comparisons between one time period and the next, it is necessary to convert nominal GDP to real GDP. That process requires dividing the rise in nominal quantities into a real component and an inflation component through the use of an appropriate price index. The overall GDP price deflator

⁴³For a detailed analysis of the sources used to estimate these sales/purchases, see "Updated Summary of NIPA Methodologies" (Bureau of Economic Analysis 2015). Boskin (2000) shows that these estimates are subject to substantial revisions, with nearly all revisions from 1959 to 1998 in the upward direction, and some of these revisions being quite large. In this journal, Landefeld, Seskin, and Fraumeni (2008) provide a very useful description of how nominal GDP and related measures are estimated from a variety of primary sources.

uses components based on the Consumer Price Index (CPI) and the Producer Price Index (PPI), requiring estimates done by the Bureau of Labor Statistics of the US Department of Labor and by the Bureau of Economic Analysis of the US Department of Commerce.

For each good and service, there are three possibilities when one compares one year with the next: (1) it is the same good or service with the same quality as in the previous period; (2) it is essentially the same good, but of a different quality; or (3) it is a wholly new good. Each category receives a different treatment in the official US statistics.

Fortunately, most goods and services fall in the first category of “no (significant) change in quality.” For those products, it is possible to collect the number of physical units sold and the total revenue. The percentage increase in revenue in excess of the percentage increase in physical volume is pure inflation, and the rest is the rise in real output. When exactly the same good is not available in the second period, the US Bureau of Labor Statistics tries to find a very similar good that does exist in the two successive periods and compares the revenue growth and physical quantity growth for that good. The BLS calls this procedure the “matched model” method.

Although much of the growth in the real value of economic output reflects substantial quality change and the introduction of wholly new products, the official procedures do not adequately reflect these sources of increased value. For products that experience quality change, the official methods tell us more about the increase in the value of inputs, in other words about the change in the cost of production, and not much about the increased value to the consumer or other ultimate user. This is true for goods as well as for services, although measuring quality improvement for services is even more difficult than it is for goods.

The government statisticians divide the period-to-period increase in total spending on each unit of product into a part due to a pure price increase (“inflation”) and a part due to an increase in quality. The part attributed to a quality increase is considered an increase in the quantity of output although, as I will explain, the method used by the BLS means that it is generally a measure of the quantity of inputs.

The Bureau of Labor Statistics is responsible for creating the Consumer Price Index and the Producer Price Indexes, as well as a number of subsidiary indexes for various categories.⁴⁴ One main difference between the CPI and

⁴⁴For a clear description of the methods of the US Bureau of Labor Statistics, see the BLS *Handbook of Methods*. Chapter 14 discusses the PPI indexes (available at www.bls.gov/opub/hom/pdf/homch14.pdf), and Chapter 17 spells out the CPI indexes (available at www.bls.gov/opub/hom/pdf/homch17.pdf).

the PPI is that many of the PPI indexes are used primarily to deflate the prices of intermediate products, rather than to deflate output for final demand. The Bureau of Economic Analysis uses those price indexes and other data to create measures of real output. These estimates are also used for measuring the output of the nonfarm business sector and are used by the Department of Commerce to calculate the GDP deflator and real GDP. The same underlying data are also used to calculate the Personal Consumption Expenditures Price Index that the Federal Reserve uses for its price stability target.⁴⁵

The key question is how the Bureau of Labor Statistics estimates the change in price when there is a change in the quality of the good or service. The BLS asks the producer of each good or service whether there has been a change in the product made by that producer. If there has been no change in the product, any change in its price is considered to be pure inflation as called for in the “matched model method.”

If a change has occurred, one approach to estimating the quality change is the “hedonic regression” method originally developed by Griliches (1961). The basic idea, which was used extensively for computers, is to regress the prices of computers in year t on a variety of the computers’ capacity and performance measures. This gives an implicit price for each of these features (if the linearity assumption of the model is correct). Applying these implicit prices to a computer model in year $t + 1$ generates a price that would apply for that computer if the values of the individual features at time t had continued to prevail.

For example, a variety of econometric studies showed that the true price of mainframe computers assessed in this way declined at an annual rate of more than 20% per year during the period from 1950 to 1980 (Chow 1967; Baily and Gordon 1989; Triplett 1989). For personal computers, Berndt, Griliches, and Rappaport (1995) found a 28% annual rate of quality-adjusted price decline during a more recent period. The lack of use of hedonic regressions in these earlier decades may be part of the explanation for Robert Solow’s (1987) comment that “you can see the computer age everywhere but in the productivity statistics.”

Hedonic regressions are used for a variety of categories in the Consumer Price Index and the Producer Price Index. In the CPI, hedonic regressions are used in categories of goods that account for about one-third of the value in the basket of goods in the Consumer Price Index, including several categories of apparel, appliances, and electronics, but the main effect of hedonic analysis on the price index is in the analysis of housing, which by itself is more than

⁴⁵A list of the price indexes used to create specific output numbers is available at Bureau of Economic Analysis (2015). For details, see also Bureau of Economic Analysis (2014).

30% of the basket of goods represented in the CPI. The Bureau of Economic Analysis incorporates these estimates, and also uses hedonic price indexes to deflate nominal output, but for only about 20% of GDP.

The use of hedonics is no doubt very difficult to apply for many of these products and services for which, unlike computers, there is not a clear list of measured technical product attributes. There is also a problem of assuming that the attributes affect willingness to pay in a linear or log-linear way. According to the government, extensions of hedonics to even more products and services is limited by the lack of detailed data and staff resources required to build and maintain the hedonic models. In this journal, Hausman (2003) discusses the limitations of hedonic pricing.

When a producer indicates that a quality change has occurred and a hedonic regression is not used, the Bureau of Labor Statistics (2014, 2015a) typically uses the “resource cost method of quality adjustment,” which is based on information about the cost of production supplied by the producer. If the producer says there has been a change in the product, the BLS asks about the “marginal cost of new input requirements that are directly tied to changes in product quality.” The rationale relied on by the BLS for this input cost as a method for defining the “quality adjustment” or, equivalently, the measure of the increased output, is described in Triplett (1983).

When the resource cost method is used, the Bureau of Labor Statistics concludes that there has been a quality improvement if and only if there is such an increase in the cost of making the product or service. The government statisticians then use the marginal cost of the product change, measured as a percentage of the previous cost of the product, to calculate a share of the price rise that is due to a quality improvement and that is therefore deemed to be an increase in the output of the product. The rest is regarded as inflation. The resource cost method can also treat a decline in production cost as evidence of a decline in quality.

This resource cost method of defining an improvement in a product or service is remarkably narrow and misleading. For the very specific case where a quality improvement is exclusively the result of adding an input, it will work. But according to this method, a pure technological innovation that makes the product or service better for the consumer doesn’t count as a product improvement unless it involves an increased cost of production! In reality, product improvements generally occur because of new ideas about how to redesign or modify an existing product or service. Those changes need not involve an increased cost of production.

Government services provide an extreme version of treating costs of inputs as equivalent to the value of outputs. Government services are valued

in the GDP at their cost, and so there is no possibility of reflecting changes in government productivity or the value created by the introduction of new government services.

Although the “resource cost method” may be the most common approach for quality adjustment, it and the hedonic procedure are not the only ones. The Bureau of Economic Analysis also uses what it calls the “quantity extrapolation method” and the “direct valuation method” for a few types of output. For example, the real quantity of bank services is derived from volume data on consumers’ deposits and loans (for discussion, see Bureau of Economic Analysis 2015).

When government statisticians deal with quality change in services, they use a variety of different methods, but none of them attempts to capture changes in the true output of the service. For some services, like legal services provided to households, the Bureau of Labor Statistics creates a price index for a variety of specific services, like writing a will, and uses that price index and total expenditure to calculate the increase in real output.

The official GDP statistics for the health care industry, which accounts for more than 17% of US GDP, focus on costs of providing various categories of health services but do not seek to capture the effect of the health products and services on the health of the patient. For example, the “output” measure for hospitals recently shifted from a day of in-patient care to an episode of hospital treatment for a particular condition. Changes in the cost per episode of treatment is the corresponding price for the Producer Price Index, which is then used to deflate expenditure to get a measure of the quantity of output. Triplett (2012, p. 17), a careful analyst of the statistical health debate, concluded that there is a “very large error in measuring output generated in the medical care sector.”

More generally, as Triplett and Bosworth (2004) note, the official data imply that productivity in the health industry, as measured by the ratio of output to the number of employee hours involved in production, declined year after year between 1987 and 2001. They conclude (p. 265) that such a decline in true productivity is unlikely, but that officially measured productivity declines because “the traditional price index procedures for handling product and service improvements do not work for most medical improvements.” More recent data show that health-sector productivity has continued to decline since 2001.

None of these measures of productivity attempt to value the improved patient outcomes. As one concrete example, when Triplett and Bosworth (2004, p. 335) wrote about the remarkable improvement in treating cataracts—from more than a week as an immobilized hospital inpatient to a

quick outpatient procedure—they questioned whether accounting for medical improvements like that would cross over the traditional “production boundary in national accounts” and asked whether “the increased value to the patient of improvement in surgery . . . belongs in national accounts if no additional charges are made.”

The Department of Commerce is experimenting with health-sector “satellite accounts” that calculate the cost of treating a patient with a particular diagnosis for a calendar year, including the cost of hospital care, physicians, and pharmaceuticals. But these accounts also do not try to capture the value of improved health outcomes. There are some research studies that attempt to measure the effect of a certain treatment on such health outcomes as quality-adjusted life years (QALYs) or disability-adjusted life years (DALYs).

For another example of the difficulties of adjusting for quality in a service, consider mutual fund management. The Bureau of Labor Statistics (2015b) has noted a substantial expansion over time in the types of funds that are available (including exchange-traded funds, fund-of-funds, long-short funds, a large number of emerging market funds, and more), but it ignores this increase in diversity of products and focuses only on the measuring output of mutual fund providers based on a percentage of all assets, concluding, “Under the current methodology, no special procedures are necessary for adjusting for the changes in the quality of portfolio management transactions” (p. 13).

To study the growth of output and productivity for individual industries, the Bureau of Labor Statistics sometimes measures real output at the industry level by the quantity of services provided. For passenger air travel, output of the industry is the number of passenger miles and productivity is defined as passenger miles per employee hour. The analysis of output “does not account for changes in service quality such as flight delays and route circuitry” (Duke and Torres 2005).

From time to time the Bureau of Labor Statistics reexamines its approach to a particular industry. When the productivity program reexamined its measure of the commercial banking industry in 2012, it revised the activities of commercial banks and raised the estimated annual output growth from 1987 to 2010 by 58%, from 2.4% a year to 3.9% a year (Royster 2012, p. 8).

My own judgment is that for most goods and services, the official estimate of quality change contains very little information about the value of the output to consumers and other final purchasers. As a result, the corresponding official measures of total real output growth are underestimates, and there is a substantial but unknown upward bias in the measure of price inflation. We don’t know what the true values are, and we don’t know how wide a margin of error there is around the official estimates.

Dealing with New Products

Although the sales of new products become immediately a part of nominal GDP, the extent to which they increase the real incomes of consumers is underestimated. Similarly, the effects of new products are not well reflected in the measures of real output and in price indexes. Moreover, the resource cost method and other government procedures for valuing changes in quality do not provide an approach to dealing with the value to consumers of new goods and services.

Instead, new products and services are not even reflected in the price indexes used to calculate real incomes and output until they represent a significant level of expenditures. They are then rotated into the sample of products used for price index calculations, and subsequent changes in their price are taken into account in the usual way. It is only at that secondary stage, sometime long after the new product has been introduced, that it affects officially measured changes in real output.

As an example to clarify how this works in practice, consider statins, the remarkable class of drugs that lowers cholesterol and reduces deaths from heart attacks and strokes. By 2003, statins were the best-selling pharmaceutical product in history and had become part of the basket of goods and services measured for the Consumer Price Index. When patents on early versions of statins then expired and generic forms became available, their prices fell. The Bureau of Labor Statistics recorded those price declines, implying a rise in real incomes. But the official statistics never estimated any value for the improvement in health that came about as a result of the introduction of statins.

To understand the magnitude of the effect of omitting the value of that single health care innovation, here is a quick history of the impact of statins. In 1994, researchers published a five-year study of 4,000-plus patients. They found that taking a statin caused a 35% reduction in cholesterol and a 42% reduction in the probability of dying of a heart attack. It didn't take long for statins to become a best-selling product with dramatic effects on cholesterol and heart attacks. According to the US Department of Health and Human Services (2011, Fig. 17, p. 26), between 1999–2002 and 2005–2008, the percentage of men aged 65–74 taking a statin doubled to about 50%. High cholesterol levels declined by more than half among men and women over age 75, and the death rate from heart disease among those over 65 fell by one-third. Grabowski et al. (2012) calculated that the combination of reduced mortality and lower hospital costs associated with heart attacks and strokes in the year 2008 alone was some \$400 billion, which was almost 3% of GDP in that year. None of this value produced by statins is included in the government's estimate of increased real income or real GDP.

This example of how statins have been treated in the national income statistics is representative of how all new products and services are treated. The value to consumers of a new good or service is ignored when the new product is at first introduced. Its price level becomes part of the Consumer Price Index when spending on that good or service is large enough to warrant inclusion. Subsequent declines in the price of the product are treated as real income gains, while price increases are part of inflationary real income losses. In short, the basic value to the consumer of the new good is completely ignored.

Ignoring what happens at the time of introduction of new products is therefore a serious further source of understating the real growth of output, incomes, and productivity. In addition, new products and services are not only valuable in themselves but are also valued by consumers because they add to the variety of available options. In an economy in which new goods and services are continually created, their omission in the current method of valuing aggregate real output makes the existing measure of real output even more of a continually increasing underestimate of true output. Hulten (2015, p. 2) summarizes decades of research on dealing with new products done by the Conference on Research in Income and Wealth with the conclusion that “the current practice for incorporating new goods are complicated but may miss much of the value of these innovations.”

The introduction of new products into the official price indexes has historically also been subject to remarkably long delays. The Boskin Commission (Boskin et al. 1996) noted that at the time of their report in 1996 there were 36 million cellular phones in the United States, but their existence had not yet been included in the Consumer Price Index. The earlier Stigler Commission (Stigler 1961) found that decade-long delays were also noted for things like room air conditioners. Autos were introduced to the Consumer Price Index only in 1940, and refrigerators in 1934. More recently, the Bureau of Labor Statistics has introduced procedures that cause new products to be rotated into the analysis more quickly, but only after they have achieved substantial scale in spending. These delays cause the price index to miss the gains from introducing the product in the first place as well as the declines in prices that often happen early in product cycles.

But these delays in the introduction of new products to the price indexes are not the key problem. Much more important is the fact that the official statistics ignore the very substantial direct benefit to consumers of new products *per se*, causing an underestimate of the rate of increase in real output and an overestimate of the corresponding rate of increase of the price index.

There is great uncertainty about the size of these potential biases. For example, the Boskin Commission (Boskin et al. 1996) was charged by the

US Senate with calculating the bias in the Consumer Price Index that is used for adjusting Social Security for changes in retirees' cost of living. The Commission considered several sources of bias in the existing Consumer Price Index, including the bias caused by changes in quality and by the omission of new products and provided estimates of each type of bias in the CPI (see also the discussion of the report in the winter 1998 issue of this journal).

But because the Boskin Commission was not able to do new research on the issue of quality change and innovation bias, it drew on existing research and on personal perceptions. For example, for "food and beverage," which accounts for 15% of the CPI, the commission members asked themselves how much a consumer would be willing to pay "for the privilege of choosing from the variety of items available in today's supermarket instead of being constrained to the much more limited variety available 30 years ago." The Boskin Commission concluded, based on pure introspection, that "a conservative estimate . . . might be 10% for food consumed at home other than produce, 20% for produce where the increased variety in winter (as well as summer farmers' markets) has been so notable, and 5% for alcoholic beverages." They used these numbers for 30 years and converted them to annual average rates of change for the 30-year period. This may be plausible, or not, but there is no real basis for believing that any of these estimates is even vaguely accurate.

Housing is the most heavily weighted component of the Consumer Price Index with a weight of nearly one-third. The Boskin Commission (Boskin et al. 1996) concluded that "a conservative estimate is that the total increase in apartment quality per square foot, including the rental value of all appliances, central air conditioning, and improved bathroom plumbing, and other amenities amounted to 10% over the past 40 years, or 0.25% per year." Maybe that is right, or maybe a better estimate would be 1% per year. There is nothing in the commission's report that helps to choose between differences of this magnitude.

In the end, the Boskin Commission concluded that the weighted average of these individual biases implied a total bias from product innovation and quality change in the annual CPI inflation rate for 1996 of 0.6 percentage points. I have no idea how much margin of error should be attached to that estimate. It served to satisfy the background political purpose for the Boskin Commission of providing a politically acceptable basis for reducing the rate of increase of Social Security benefits.

A formal analytic approach to the problem of valuing new products was developed by Hausman (1996, 2003). He showed how the value to consumers of a single new product could be measured by estimating the value of

introducing a new brand of breakfast cereal—specifically Apple-Cinnamon Cheerios. His approach, following the theory presented by Hicks (1940), was to estimate the “virtual price,” that is, the price that would prevail when the good is just introduced at zero quantity. The consumer gains an amount of real income when the good is introduced, implied by the decline in its price from the virtual price to the actual market price. He concluded that the Consumer Price Index component for cereals may be overstated by about 20% because of its neglect of new cereal brands. The Hausman estimates were controversial, but if the magnitude is even roughly indicative of the overstatement of the Consumer Price Index from a failure to reflect the introduction of new varieties of cereal brands, then surely the overstatement of the Consumer Price Index and the understatement of real income that result from failing to take into account new products like statins and new anticancer drugs must be substantially larger.

Broda and Weinstein (2010) and Redding and Weinstein (2016) extend the Hausman (1996) approach and present a new method for valuing new products as well as the value to consumers of changes in product quality. They analyze a very large set of data on bar-coded package goods for which prices and quantities are available over time. By studying these data in the framework of a demand system based on constant-elasticity-of-substitution utility functions, they find that conventional price indexes overstate inflation for this set of goods by as much as 5 percentage points because the conventional measure ignores quality and new goods biases. Of course, this method is limited to goods and services for which the bar-coded price and quantity data are available and requires accepting a specific theoretical demand specification for these products. But as the availability of data on prices and quantity grows, it provides a starting point for improving the overall measurement of consumer prices and the corresponding estimates of real income.

The creation of new products also means an increased variety of choice, a form of quality improvement in itself, as Hausman (1996) noted. The value to consumers of access to an increased variety of options, which allows individuals to make choices that conform to their personal taste, can be substantial. Coyle (2014) noted that in the 30 years after 1970, the number of commonly available television channels rose from 5 to 185, and the number of soft drink brands climbed from 20 to 87.

The failure to take new products into account in a way that reflects their value to consumers may be an even greater distortion in the estimate of real growth than the failure to reflect changes in the quality of individual goods and services. At present, there is no way to know.

Productivity Change and Its Recent Slowdown

Labor productivity is defined as the ratio of real output to the number of hours worked by all employed persons. The Bureau of Labor Statistics estimates labor productivity for the nonfarm business sector, as well as for some parts of that sector, using output estimates provided by the Bureau of Economic Analysis.⁴⁶

The key problem in measuring labor productivity is in the numerator—that is, in measuring output. The failure to measure quality changes adequately and to incorporate the value of new products means that true output has grown faster than measured output and therefore that the pace of productivity growth has been underestimated. This problem is particularly difficult in service industries. Bosworth and Triplett (2000, p. 6; Triplett and Bosworth 2004, p. 331) note that the official data imply that productivity has *declined* in several major service industries—including health care, hotels, education, entertainment, and recreation—and concluded that this apparent decline was “unlikely” and probably reflected measurement problems.

While the understatement of productivity growth is a chronic problem, there has been a sharp decline in the officially measured rate of productivity growth in the last decade. That sharp decline remains a puzzle that is yet to be resolved, as Syverson discusses in this issue. His work, along with papers by Fernald (2014) and Byrne, Fernald, and Reinsdorf (2016) show that the recent productivity slowdown cannot be attributed to the effects of the recession of 2008–2009, to changes in the labor force demographics in recent years, or to the growth of unmeasured internet services. One possible explanation of the recent downturn in productivity growth may be that the unusually rapid increase in the productivity growth in the prior few years was an anomaly and the recent decline is just a return to earlier productivity patterns.

A further hypothesis for explaining the recent downturn in productivity growth that has not yet been fully explored involves the mismeasurement of official estimates of output and productivity. Any attempt to explain the recent decline in the estimated productivity growth rate must attempt to understand not just the aggregate behavior for the nonfarm business sector as a whole but also what happened at the disaggregated level. (Official estimates of productivity by industry are available from the Bureau of Labor Statistics, 1987–2015, “Industry Productivity,” although it should be noted that the

⁴⁶In contrast, multifactor productivity is the ratio of real output to a combination of labor and capital input services. It is intended to measure the increase in output that is not attributable to either labor inputs or capital inputs. A good deal of research has been devoted to the very difficult problem of measuring the input of capital services and to the correct way to combine labor and capital inputs. Here, I will sidestep these issues by focusing on labor productivity.

overall productivity measure is not calculated by combining the individual industry numbers but is estimated separately based on a measure of real value added.)

The recent decline in the official measure of overall labor productivity growth in the nonfarm business sector reflects an enormous diversity of changes of productivity in specific industry groups. For the nonfarm business sector as a whole, the rate of productivity growth fell from 3.2% a year in the decade from 1995 to 2004 to just 1.5% in the decade from 2004 to 2013. The decline of 1.7 percentage points in the overall productivity change reflects an enormous range of changes in various industries. Even if attention is limited to the relatively aggregate three-digit level, the official productivity data show that productivity in apparel manufacturing went from annual growth at 1% in the earlier decade to an annual productivity decline of 5% in the later period, a drop of 6 percentage points. For manufacturing of computers and electronic products, productivity growth fell from a 15% annual rate to a 4% annual rate, a fall of 11 percentage points. Some industries experienced faster productivity growth, with productivity in the manufacturing of wood products increasing from a 2% annual rise in the early period to a 2.4% rise in the later period.

The differences are even greater at a more disaggregated level. At the four-digit level, for example, productivity growth increased by 5 percentage points annually for radio and TV broadcasting but declined by 18% for semiconductors and electronic components. The deflation of output for disaggregated industries is even harder than for the economy as a whole because nominal outputs must be deflated by quality-adjusted prices for the disaggregated industries (Dennison 1989).

It would be intriguing, although difficult, to explore how or whether productivity differences across industries might be correlated with the problems of dealing with product change and the introduction of new goods and services in those industries.

Using Our Imperfect Data

What can be learned from the imperfect measures of real output and from the corresponding overstatement of price inflation? How should our understanding of the mismeasurement affect the making of monetary and fiscal policies?

Assessing Cyclical Economic Conditions. Consider first the assessment of short-run business cycle conditions. Policymakers and financial markets often focus on short-term fluctuations of real GDP as an indication of the state of the business cycle. Although measuring the size of fluctuations of real GDP is flawed by the difficulty of dealing with new products and

quality changes, the official measure of real GDP fluctuations can in principle capture the short-term up or down changes in the pace of economic activity. Of course, it is important to recognize the substantial uncertainty about the estimated short-run fluctuations in GDP and the subsequent revisions.⁴⁷

But it is interesting to note that when the Business Cycle Dating Committee of the National Bureau of Economic Research (NBER) meets to consider appropriate dates for the start and end of a recession, it places relatively little emphasis on GDP. Contrary to popular belief, the NBER committee has never used two quarters of decline in real GDP as its definition of a recession. Instead, it has traditionally looked at employment, industrial production, wholesale-retail sales, as well as real income. In recent years, the NBER committee has also looked at monthly GDP when Macro Advisers began creating monthly estimates of GDP.

All data involve problems of interpretation in judging the state of economic activity, but employment, industrial production, and nominal sales are relatively free from the problem of quality adjustment and price measurement that affect measures of real GDP. Employment data are available monthly with substantial detail based on a large survey of employers. Industrial production is estimated by the Federal Reserve based primarily on data on physical production (such as tons of steel and barrels of oil) obtained from trade associations and government agencies, supplemented when necessary with data on production-worker hours and for some high-tech products by using nominal output and a price index (for details, see the Federal Reserve Board data at <https://www.federalreserve.gov/releases/G17/>). These measures of industrial production as well as wholesale-retail sales deal with economic activity without having to impute value in large amounts, as must be done for the services of owner-occupied homes that are involved in the estimate of GDP.

Assessing Longer-Term Growth and Inflation. For the longer term, the official measures of changes in real output are misleading because they essentially ignore the value created by the introduction of new goods and services and underestimate changes in the quality of these products. It follows therefore that “true” real output is growing faster than the official estimates

⁴⁷The Federal Reserve Banks of New York and Atlanta have recently begun using official data to produce preliminary estimates of changes in real GDP even before the corresponding quarter is over, but with some variability in results. In April 2016, the New York Federal Reserve estimated that real GDP increased by 1.1% in the recently completed first quarter of 2016, while the Atlanta Federal Reserve estimated that the increase in the same quarter was only 0.1%.

imply and that the corresponding “true” GDP price index is rising more slowly than the official one—or is actually declining.

The economics profession should educate the general public and the policy officials that “true” real incomes are rising faster than the official data imply. We can reassure people that it is very unlikely that the real incomes of future generations will be lower than real incomes today. Even if the future will not see the “epochal innovations” of the type that Kuznets (1971) referred to or such fundamental changes as electricity and indoor plumbing that caused jumps in living standards (as emphasized by Gordon 2016), current and future generations can continue to experience rising real incomes due to technological changes, improvements in education, and increases in health care technology.

One can only speculate about whether the bias in the officially measured pace of real output change is greater now than in the past. One reason to think that the gap between true output growth and measured growth is greater now than in the past is that services now represent about 70% of private value added, up from about 50% of private value added back in 1950, and the degree of underestimation of quality change and product innovation may be greater for services. Within services, health occupies a larger share of output—and quality improvements there may be greater than in other parts of the service sector. The internet and services through the internet have become much more important and are also harder to measure.

Poverty and Distribution. Trends in the overestimation of inflation and therefore in the underestimation of real incomes may vary among demographic groups and income groups because of differences in the mix of goods and services consumed by these different groups. For example, are the goods and services bought by older people improving relatively faster than the goods and services bought by younger households? Health care is an obvious example, although most of the consumption of health care by the elderly is financed by government transfers.

Implications for Fiscal and Monetary Policy. Policy issues that depend on nominal measures of output are unaffected by the problems discussed in this essay. The most obvious of these is the ratio of debt to GDP, since both the numerator and the denominator are nominal values. Similarly, the rate of change of the debt-to-GDP ratio depends only on the nominal value of the annual deficit and the annual rate of nominal GDP growth. If the debt-to-GDP ratio is not on an explosive path, its long-run equilibrium value is equal to the annual nominal deficit ratio divided by the rate of nominal GDP growth.

The evidence that the true inflation rate is less than the measured inflation rate may imply that the true inflation rate is now less than zero. Fortunately, this does not imply that the US economy is experiencing the traditional problem of debt deflation (Fisher 1933), which occurs when a declining price level reduces aggregate demand by increasing the value of household debt relative to current incomes. The traditional problem of debt deflation does not arise under current conditions because the nominal value of wage income is not declining and the real monthly wage is rising more rapidly.

Overestimating the true rate of inflation does imply that the real rate of interest is higher than the conventionally measured rate. If households recognize that their dollars will buy relatively more in the future, this could alter the household saving rate—either increasing saving in response to the greater reward for saving or decreasing saving because a given volume of assets will buy more in the future, depending on whether substitution or income effects dominate. Because many factors affect the household saving rate, it is not clear which of these effects now dominates. Uncertainty about the true rate of inflation should affect the optimal monetary policy. There seems little point in having a precise inflation target when the true rate of inflation is measured with a great deal of uncertainty. The goal of price stability also takes on a new meaning if true inflation is substantially negative while measured inflation is low but positive. Would it be better to have a target range for measured inflation as the Federal Reserve does now? Or to have a target range for measured inflation that is higher and further from the zero bound, thus leaving more room for larger changes in nominal interest rates while recognizing that the actual inflation rate is lower than the officially measured one? Or to restate the inflation goal of monetary policy as reacting when there is a rapid movement in measured inflation either up or down?

The underestimation of real growth has affected Federal Reserve decision making in the past. Back in 1996, Fed chairman Alan Greenspan persuaded members of the Federal Open Market Committee that the official data underestimated productivity growth, so maintaining strong demand would not cause a rise in inflation and there was no reason to raise interest rates (Mallaby 2016). In the last few years, the perception of slow real growth is often mentioned in support of a Federal Reserve policy of exceptionally low interest rates, but if real growth rates are actually higher (or if real growth rates have not dipped as much as the official statistics seem to show), then the Fed's policy of ultra-low interest rates has been providing little gain while contributing to certain risks of potential financial instability.

A great deal of effort and talent has been applied over past decades to the measurement of real income and inflation. These problems are extremely

difficult. In my judgment, they are far from being resolved, and as a result, substantial errors of unknown size remain in our ability to measure both real output and inflation. It is important for economists to recognize the limits of our knowledge and to adjust public statements and policies to what we can know.

I am grateful for extensive help with this paper to Anna Stansbury and for comments on earlier drafts from Katherine Abraham, Graham Allison, Michael Boskin, Erica Groshen, Jim Stock, David Weinstein, members of the staffs of the Council of Economic Advisers, the Bureau of Labor Statistics (especially David Friedman), and the Bureau of Economic Analysis, as well as participants at a meeting of the Group of Thirty and at a Harvard seminar.

References

Baily, Martin, and Robert J. Gordon. 1989. "Measurement Issues, the Productivity Slowdown and the Explosion of Computer Power." CEPR Discussion Paper 305. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=305.

Berndt, Ernst R., Zvi Griliches, and Neal J. Rappaport. 1995. "Econometric Estimates of Price Indexes for Personal Computers in the 1990s." *Journal of Econometrics* 68 (1): 243–68.

Boskin, Michael J. 2000. "Economic Measurement: Progress and Challenges." *American Economic Review* 90 (2): 247–52.

Boskin, Michael J., Ellen R. Dulberger, Zvi Griliches, Robert J. Gordon, and Dale Jorgensen. 1996. *Toward a More Accurate Measure of the Cost of Living*. Final report to the US Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index. <https://www.ssa.gov/history/reports/boskinrpt.html>.

Bosworth, Barry, and Jack E. Triplett. 2000. "Numbers Matter: The U.S. Statistical System and a Rapidly Changing Economy." Brookings Policy Brief 63.

Broda, Christian, and David E. Weinstein. 2010. "Product Creation and Destruction: Evidence and Price Implications." *American Economic Review* 100 (3): 691–723.

Bureau of Economic Analysis, US Department of Commerce. 2019. *NIPA Handbook: Concepts and Methods of the U.S. National Income and Product Accounts*. Washington, DC. <https://www.bea.gov/media/5736>.

_____. 2015. "Updated Summary of NIPA Methodologies." *Survey of Current Business* 95 (11). <https://apps.bea.gov/scb/toc/1115cont.htm>.

Bureau of Labor Statistics, US Department of Labor. 2016. "Industry Productivity" (1987–2015). LCP (Labor, Productivity, and Costs) Database. <https://www.bls.gov/lpc/data.htm>.

_____. 2016. "Major Sector Productivity and Costs" (1987–2015). LCP (Labor, Productivity, and Costs) Database. <https://www.bls.gov/lpc/data.htm>.

_____. 2011. "Producer Prices." In *Handbook of Methods*, chapter 14. <http://www.bls.gov/opub/hom/pdf/homch14.pdf>.

_____. 2018. "The Consumer Price Index." In *Handbook of Methods*, chapter 17. <http://www.bls.gov/opub/hom/pdf/homch17.pdf>.

_____. 2014. "Quality Adjustment in the Producer Price Index." <http://www.bls.gov/ppi/qualityadjustment.pdf>.

_____. 2015a. "Hedonic Models in the Producer Price Index (PPI)." <http://www.bls.gov/ppi/ppicomqa.htm>. First published 2011.

_____. 2015b. "Producer Price Indexes Introduced for Portfolio Management—NAICS 523920 and Investment Advice—523920." <https://www.bls.gov/ppi/ppiinvestadvice.htm>.

Byrne, David M., John G. Fernald, and Marshall B. Reinsdorf. 2016. "Does the United States Have a Productivity Slowdown or a Measurement Problem?" *Brookings Papers on Economic Activity* 2016 (Spring): 109–57. <http://www.jstor.org/stable/43869021>.

Chow, Gregory C. 1967. "Technological Change and the Demand for Computers." *American Economic Review* 57 (5): 1117–30.

Council of Economic Advisers, Chair. 2015. *Economic Report of the President*. Transmitted to the US Congress February 2015. Washington, DC: US Government Printing Office.

Coyle, Diane. 2014. *GDP: A Brief But Affectionate History*. Princeton, NJ: Princeton University Press.

Dennison, Edward F. 1989. *Estimates of Productivity Change by Industry: An Evaluation and an Alternative*. Washington, DC: Brookings Institution Press.

Dugan, Andrew. 2016. "U.S. Economic Confidence Index Flat at -10." Gallup, 18 October. <http://www.gallup.com/poll/196526/economic-confidence-index-flat.aspx>.

Duke, John, and Victor Torres. 2005. "Multifactor Productivity Change in the Air Transportation Industry." *Monthly Labor Review* 128 (3): 32–45.

Federal Reserve Bank of Atlanta. 2016. "GDP Now." <https://www.frbatlanta.org/cqer/research/gdpnow.aspx?panel=1>.

Federal Reserve Bank of New York. 2016. "Nowcasting Report: April 8, 2016." Staff Nowcast. https://www.newyorkfed.org/medialibrary/media/research/policy/nowcast/nowcast_2016_0408.pdf?la=en.

Federal Reserve System, Board of Governors. 2014. *Report on the Economic Well-Being of U.S. Households in 2013*. <https://www.federalreserve.gov/econresdata/2013-report-economic-well-being-us-households-201407.pdf>.

Federal Reserve System, Board of Governors. "Industrial Production and Capacity Utilization—G.17." <http://www.federalreserve.gov/releases/g17>.

Fernald, John G. 2014. "Productivity and Potential Output Before, During and After the Great Recession." NBER Working Paper 20248.

Fisher, Irving. 1933. "The Debt-Deflation Theory of Great Depressions." *Econometrica* 1 (4): 337–57.

Frazis, Henry, and Jay Stewart. 2011. "How Does Nonmarket Production Affect Measured Earnings Inequality." *Journal of Population Economics* 24 (1): 3–22.

Gordon, Robert J. 2016. *The Rise and Fall of American Growth: The US Standard of Living since the Civil War*. Princeton, NJ: Princeton University Press.

Grabowski, David C., Darius N. Lakdawalla, Dana P. Goldman, Michael Eber, Larry Z. Liu, Tamer Abdelgawad, Andreas Kuznik, Michael E. Chernew, and Tomas Philipson. 2012. "The Large Social Value Resulting from Use of Statins Warrants Steps to Improve Adherence and Broader Treatment." *Health Affairs* 31 (10): 2276–85.

Griliches, Zvi. 1961. "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change." In *The Price Statistics of the Federal Government: Review, Appraisal, and Recommendations*, General Series 73, 173–96. Report to the Offices of Statistical Standards Bureau of the Budget prepared by the Price Statistics Review Committee of the National Bureau of Economic Research.

Griliches, Zvi, ed. 1992. *Output Measurement in the Service Sectors*. National Bureau of Economic Research Studies in Income and Wealth. Chicago: University of Chicago Press.

Hausman, Jerry A. 1996. "Valuation of New Goods Under Perfect and Imperfect Competition." In *The Economics of New Goods*, edited by Timothy F. Bresnahan and Robert J. Gordon, 207–48. National Bureau of Economic Research. Chicago: University of Chicago Press.

Hausman, Jerry A. 1999. "Cellular Telephone, New Products, and the CPI." *Journal of Business & Economic Statistics* 17 (2): 188–94.

Hausman, Jerry A. 2003. "Sources of Bias and Solutions to Bias in the Consumer Price Index." *Journal of Economic Perspectives* 17 (1): 23–44.

Hicks, John R. 1940. "The Valuation of the Social Income." *Economica* 7 (26): 105–24.

Hulten, Charles R. 2015. "Measuring the Economy of the 21st Century." *NBER Reporter* 2015 (4): 1–7.

Kuznets, Simon. 1934. "National Income, 1929–1932." Bulletin 49, National Bureau of Economic Research (7 July).

Kuznets, Simon. 1971. "Modern Economic Growth: Findings and Reflections." Sveriges Riksbank Prize in Economic Sciences Lecture to the Memory of Alfred Nobel, 11 December. <https://www.nobelprize.org/prizes/economic-sciences/1971/kuznets/lecture/>.

Kuznets, Simon, Lillian Epstein, and Elizabeth Jenks. 1941. *National Income and Its Composition, 1919–1938*, vol. 1. National Bureau of Economic Research.

Landefeld, J. Steven, Eugene P. Seskin, and Barbara M. Fraumeni. 2008. "Taking the Pulse of the Economy: Measuring GDP." *Journal of Economic Perspectives* 22 (2): 193–216.

Long, Heather. 2016. "56% of Americans Think Their Kids Will Be Worse Off." *CNN Money* (28 January). <http://money.cnn.com/2016/01/28/news/economy/donald-trump-bernie-sanders-us-economy>.

Mallaby, Sebastian. 2016. *The Man Who Knew: The Life and Times of Alan Greenspan*. Bloomsbury.

Marshall, Alfred. 1925 (1887). "Remedies for Fluctuations of General Prices." In *Memorials of Alfred Marshall*, edited by A.C. Pigou. London: MacMillan.

Mitchell, Wesley Clair, Wilford Isbell King, and Frederick R. Macaulay. 1921. "Introduction." In *Income in the United States: Its Amount and Distribution, 1909–1919, Volume 1: Summary*. New York: National Bureau of Economic Research.

National Research Council. 2002. *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes*. Edited by Charles Schultze and Christopher Mackie. Washington, DC: National Academy Press.

Nordhaus, William D. 1997. "Traditional Productivity Estimates Are Asleep at the (Technological) Switch." *Economic Journal* 107 (444): 1548–59.

Redding, Stephen J., and David E. Weinstein. 2016. "A Unified Approach to Estimating Demand and Welfare." NBER Working Paper 22479. <http://www.princeton.edu/~reddings/papers/NBER-WP-22479.pdf>.

Royster, Sara E. 2012. "Improved Measures of Commercial Banking Output and Productivity." *Monthly Labor Review* 135 (7): 3–12.

Sidgwick, Henry. 1883. *The Principles of Political Economy*. London: Macmillan and Co.

Solow, Robert. 1987. "We'd Better Watch Out." *New York Times Book Review* (12 July): 36.

Stigler, George, ed. 1961. *The Price Statistics of the Federal Government: Review, Appraisal, and Recommendations*, General Series 73. Report to the Office of Statistical Standards Bureau of the Budget prepared by the Price Statistics Review Committee of the National Bureau of Economic Research.

Triplett, Jack E. 1983. "Concepts of Quality in Input and Output Price Measures: A Resolution of the User-Value Resource-Cost Debate." In *The U.S. National Income and Product Accounts: Selected Topics*, edited by Murray F. Foss, chapter 5. Chicago: University of Chicago Press.

———. 1989. "Price and Technological Change in a Capital Good: A Survey of Research on Computers." In *Technology and Capital Formation*, edited by Dale W. Jorgenson and Ralph Landau, 127–213. Cambridge, MA: MIT Press.

———. 2012. "Health System Productivity." In *The Oxford Handbook of Health Economics*, edited by Sherry Glied and Peter C. Smith, chapter 30. Oxford, England: Oxford University Press.

Triplett, Jack E., and Barry P. Bosworth. 2004. *Productivity in the US Services Sector: New Sources of Economic Growth*. Washington, DC: Brookings Institution Press.

US Department of Health and Human Services. 2011. “Health, United States, 2010: With Special Feature on Death and Dying.” <https://www.cdc.gov/nchs/data/hus/hus10.pdf>.

Challenges to Mismeasurement Explanations for the US Productivity Slowdown

Chad Syverson

J. Baum Harris Professor of Economics, University of Chicago Booth School of Business

The flow and ebb of US productivity growth since World War II is commonly divided into four periods: 1947–1973, 1974–1994, 1995–2004, and 2004–2015. After labor productivity growth averaged 2.7% per year from 1947 to 1973, it fell in a much studied but still debated slowdown to 1.5% per year over 1974–1994. Another fast/slow cycle has followed. Productivity growth rose to a trajectory of 2.8% average annual growth sustained over 1995–2004. But since then, the US economy has been experiencing a slowdown in measured labor productivity growth. From 2005 through 2015, labor productivity growth has averaged 1.3% per year (as measured by the nonfarm private business labor productivity series compiled by the US Bureau of Labor Statistics).

This slowdown is statistically and economically significant. A t -test comparing average quarterly labor productivity growth rates over 1995–2004 to those for 2005–2015 rejects equality with a p -value of 0.008. If the annualized 1.5 percentage point drop in labor productivity growth were to be sustained for 25 years, it would compound to an almost 50% difference in income per capita.

The productivity slowdown does not appear to be due to cyclical phenomena. Fernald (2014a) shows that the slowdown started before the onset of the Great Recession and is not tied to “bubble economy” phenomena in housing or finance. This work, along with the analysis in Byrne, Oliner, and Sichel (2013), ties the slowdown to a reversal of the productivity accelerations in the manufacturing and utilization of information and communication technologies that drove the more rapid pace of productivity from 1995 to 2004. While one cannot rule out persistent, less-direct channels through which the Great

Reprinted from the *Journal of Economic Perspectives*, Spring 2017, with the permission of the publisher, the American Economic Association.

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.31.2.165doi=10.1257/jep.31.2.165>.

Recession might have long-lived influences on productivity growth, it is clear that measured labor productivity in the United States has not awakened from its slowdown as the Great Recession recedes.

The debate about the causes of the productivity slowdown is ongoing. Gordon (2016) points to multiple possible explanations and ties the current slowdown to the one in 1974–1994, viewing the 1995–2004 acceleration as a one-off aberration. Cowen (2011) shares these views and enumerates multiple reasons why innovation—at least the kind that leads to changes in measured productivity and income—may slow. Tarullo (2014) suggests that the slowdown in US business dynamism documented by Decker, Haltiwanger, Jarmin, and Miranda (2014) and Davis and Haltiwanger (2014) may have a role. Some have argued that there are reasons to be optimistic that the slowdown may reverse itself. Baily, Manyika, and Gupta (2013) point to potential innovation opportunities in multiple sectors. Syverson (2013) notes that the productivity growth from electrification and the internal combustion engine—a prior diffusion of a general-purpose technology—came in multiple waves, implying that the 1995–2004 acceleration need not be a one-time event.

However, these arguments all accept that the measured decline in productivity growth is meaningful. A separate set of explanations for the slowdown in measured productivity put forward by several parties is that it is substantially illusory (for example, Brynjolfsson and McAfee, 2011, 2014; Mokyr 2014; Alloway 2015; Byrne, Oliner, and Sichel 2015; Feldstein 2015; Hatzius and Dawsey 2015; Smith 2015). The theme of these arguments is that true productivity growth since 2004 has not slowed as much as official statistics may suggest—and perhaps productivity growth has even accelerated—but that due to measurement problems, the new and better products of the past decade are not being captured in official productivity metrics.

There is a *prima facie* case for this assertion, which for brevity I refer to as the “mismeasurement hypothesis.” Many of the fastest-diffusing technologies since 2004—like smartphones, online social networks, and downloadable media—involve consumption of products that are time intensive but do not impose a large direct monetary cost on consumers. If one considers the total expenditure on such products to be both the monetary price and the value of time spent consuming them, a revealed preference argument would suggest they deliver substantial utility (Becker 1965). At the same time, the fact that these new products are not particularly expensive (at least relative to consumers’ supposed interest in them) could result in a relatively modest portion of their delivered consumption benefit to be reflected in GDP.

This mismeasurement hypothesis could take one of two related forms. One possibility is that a smaller share of the utility that these products provide

is embodied in their prices than was the case for products made before 2004. If this were true, measured output growth would slow even as growth of total surplus continued apace. The second possibility is that if the price deflators of these new technology products are rising too fast (or falling too slowly) relative to their pre-2004 changes, the result would be that quantity growth as backed out from nominal sales is understated.⁴⁸

In this study, I explore the quantitative plausibility of the mismeasurement hypothesis. One fact dominates the discussion: Had the measured productivity slowdown not happened, measured GDP in 2015 would have been, conservatively, \$3 trillion (17%) higher than it was. This is \$9,300 for every person or \$24,100 for every household in the United States. For the mismeasurement hypothesis to explain the productivity slowdown, the losses in measured incremental gains from the new technologies would need to be at or around this level. Thus, to explain even a substantial fraction of the productivity slowdown, current GDP measures must be missing hundreds of billions of dollars of incremental output (and moreover with no accompanying employment growth).

I start with a computation of the missing output lost to the productivity slowdown. I then turn to discussion of four patterns in the data, each looking at the mismeasurement hypothesis from different directions, which pose challenges for the hypothesis.

First, the productivity slowdown is not unique to the United States. It has occurred with similar timing across at least two dozen other advanced economies. However, the magnitude of the productivity slowdown across countries (of which there is nontrivial variation) is unrelated to the relative size of information and communication technologies (ICT) in the country's economy, whether this "ICT intensity" is measured in consumption or production terms.

Second, a research literature has attempted to measure the consumer surplus of the internet. These efforts are based on the notion that many of the

⁴⁸These issues have arisen before. Diewert and Fox (1999) discuss related productivity measurement problems in the context of an earlier slowdown, arguing that there were several plausible sources of mismeasurement. The price-deflator-based interpretation of the measurement problem evokes the Boskin Commission report (US Congress 1996), which argued that the Consumer Price Index methodology at the time overstated inflation and therefore understated growth. Many of the commission's suggested changes, including those specifically aimed at better measurement of new products and technologies, were implemented before 2004 (Klenow 2003). The issues raised by the Boskin Commission report were discussed in a six-paper symposium on "Measuring the CPI" in the winter 1998 issue of this journal, and a follow-up report by the National Academy of Sciences was discussed in a three-paper symposium on the "Consumer Price Index" in the winter 2003 issue.

newer technologies that could create large surplus with little revenue require internet access, which makes purchase and use of internet access a metric for the gains from such technologies. However, most of the estimates of the value of internet-linked technologies are at least an order of magnitude smaller than the trillions of dollars of measured output lost to the productivity slowdown. As I will discuss, even the largest estimate, which explicitly accounts for the time people spend online and is computed with very generous assumptions about the value of that time, totals only about one-third of the missing output.

Third, if the mismeasurement hypothesis were to account entirely (or almost so) for the productivity slowdown, and if the source of this mismeasurement is predominantly in certain industries that make and service digital and information and communication technologies, then the implied change in real revenues of these industries would be five times their measured revenue change. Incremental real value added would have been six times the observed change, and true labor productivity in these industries would have risen 363% over 11 years.

Fourth, gross domestic income (GDI) and gross domestic product (GDP) are conceptually equivalent, but because they are computed with different source data, they are not actually equal. Since 2004, GDI has outstripped GDP by an average of 0.4% of GDP per year. This pattern is consistent with workers being paid to produce goods that are being given away for free or sold at steep discounts, which is consistent with the mechanism behind the mismeasurement hypothesis. However, I show that GDI began to be larger than GDP in 1998—several years before the productivity slowdown and, indeed, in the midst of a well-documented productivity acceleration. Additionally, a breakdown of GDI by income type shows that GDI growth over the period has been driven by historically high capital income (like corporate profits), while labor income has actually fallen. This is opposite the implication of a “workers paid to make products sold free” story.

In isolation, none of these four patterns is dispositive. But taken together, they challenge the ability of the mismeasurement hypothesis to explain a substantial part of the productivity slowdown.

Calculating the Missing Output

Whether the mismeasurement of productivity hypothesis is presumed to act through output gains disproportionately flowing into consumer surplus rather than GDP or through incorrect price deflators, the implication is the same: US consumers benefited from this missing output, but it just was not reflected in measured GDP. Any evaluation of the hypothesis needs to put estimates of

productivity mismeasurement in the context of measures of this hypothetically missing output.

I first compute the implied lost output due to the productivity slowdown. Using quarterly labor productivity data from the US Bureau of Labor Statistics for the entire nonfarm business sector, I calculate average quarterly productivity growth over four post–World War II periods: 1947–1973, 1974–1994, 1995–2004, and 2005–2015 (period averages are inclusive of endpoint years). Past research has shown that average productivity growth has inflection points at or around the transitions between these periods, and work on both the most recent and prior productivity slowdowns has used these periods (for example, Byrne, Oliner, and Sichel 2013). **Table 1** shows average productivity growth rates along with their annualized values for each period. As is clear in the table, measured labor productivity growth after 2004 fell by more than half from its 1995–2004 average.⁴⁹

Labor productivity is defined as the ratio of real output to labor inputs, so it is straightforward to compute what counterfactual output would have been after 2004 had productivity growth not slowed. The drop in average quarterly labor productivity growth between 1995–2004 and 2005–2015 is 0.395 percentage points ($= 0.712 - 0.317$). Thus, counterfactual output in 2015 would

Table 1. Average Quarterly Labor Productivity (LP) Growth by Period

Period	Average Quarterly LP Growth (%)	Annualized LP Growth (%)
1947–1973	0.681	2.73
1974–1994	0.386	1.54
1995–2004	0.712	2.85
2005–2015	0.317	1.27

Notes: These values are taken from the Bureau of Labor Statistics nonfarm private industry labor productivity growth series. Annualized growth values are simply four times quarterly growth. Notably, these sectors might vary in their inherent “measurability.” Total factor productivity growth also slowed. The Bureau of Labor Statistics measure of multifactor productivity fell from 1.4% per year during 1995–2004 to 0.5% per year over 2005–2015. The utilization-corrected total factor productivity measures of Fernald (2014b) also saw similar decelerations, by 2.5% per year in the equipment- and consumer durables-producing sectors and 1.1% per year for makers of other outputs.

⁴⁹Related productivity measures testify to the spread and depth of the slowdown. Sector-specific labor productivity growth slowed over the same period for each of the six two-digit NAICS industries with available data (mining, utilities, manufacturing, wholesale, retail, and accommodation and food services).

have been 19% higher ($1.0039544 = 1.189$) than observed output in that period. Note that this exercise does not change labor inputs. Counterfactual output still reflects the observed movements in labor inputs over the period, like the considerable decline during the Great Recession. This exercise therefore does not assume away the employment downturn of the slowdown period.⁵⁰

Nominal GDP in 2015 was \$18.037 trillion. If I apply the counterfactual extra productivity growth of 19% to this value, the amount of output “lost” due to the productivity slowdown is \$3.43 trillion per year.⁵¹

However, it is not immediately obvious if GDP is the correct base to which to apply the counterfactual growth rate. The Bureau of Labor Statistics labor productivity series that I use here applies to nonfarm business activity, which excludes farming, government, nonprofits, and paid employees of private households. The reason given is that the outputs of these sectors in GDP “are based largely on the incomes of input factors. In other words, the measure is constructed by making an implicit assumption of negligible productivity change” (www.bls.gov/lpc/faqs.htm). The value of owner-occupied dwellings is left out “because this sector lacks a measure of the hours homeowners spend maintaining their home.” Together, these factors jointly account for about one-quarter of GDP. If labor productivity growth in the excluded activities didn’t slow as much as in nonfarm business productivity growth, then the “lost” output could be smaller than \$3.43 trillion per year; conversely, if productivity in the excluded activities slowed more, then the “lost” output could be larger. As long as productivity growth did not actually

⁵⁰An implication of the mismeasurement hypothesis is that the reported output deflator does not reflect true price changes and should have grown more slowly than what was measured. It is therefore instructive to compare the average growth rates of the implicit price deflator for the Bureau of Labor Statistics productivity series in the 1995–2004 and 2005–15 periods. The deflator grew an average of 0.36% per quarter from 1995–2004 and 0.41% per quarter from 2005–2015. Compounded over the 44 quarters of the latter period, the deflator grew a cumulative 2.3% more than had it remained at its earlier trajectory. To the extent that this acceleration might reflect real output mismeasurement (and the fact that it did accelerate does not imply that it shouldn’t have), it would explain only about one-eighth of the measured slowdown.

⁵¹The calculations here and throughout this paper use 2015 as an endpoint because several of the data sources I use extend only through that year. The implied “lost” output would be even larger than the reported values if I used the labor productivity data through 2016 (the latest available numbers as of this writing). This is for two reasons. First, average labor productivity growth during 2016 was even slower than the 2005–15 average. Second, the slowdown would be compounded over another year of GDP growth. Conducting similar calculations to those above using the 2016 data imply values of lost output that are 14% larger than those reported here.

accelerate in these excluded sectors—which seems a fair assumption—a very conservative estimate of lost output would apply the 19% slowdown only to the three-fourths of GDP that the labor productivity series covers directly. This lower bound implies at least \$2.57 trillion of lost output.

Some additional data can refine this lower bound estimate. First, the Bureau of Labor Statistics does compute a productivity series that adds the farming sector (which accounts for about 1% of GDP) to the set of covered industries. This series experienced an even larger productivity slowdown than the nonfarm business series, falling from an average growth per quarter of 0.741% over 1995–2004 to 0.310% for 2005–2015. This implies a larger amount of “missing” output—\$3.80 trillion applied to GDP or a lower bound of about \$2.89 trillion when applied only to the directly covered sectors. Second, I combined an unpublished Bureau of Labor Statistics series of total economy aggregate hours through 2015 with the real GDP index from the Bureau of Economic Analysis to compute a total economy labor productivity measure.⁵² This metric indicates a drop in productivity growth between 1995–2004 and 2005–2015 of 0.369 percentage points per quarter. Applying this to all of GDP (which, here, the productivity metric spans) implies lost output due to the productivity slowdown of \$3.21 trillion per year.

Thus, the amount of output lost to the productivity slowdown ranges somewhere between \$2.57 trillion and \$3.80 trillion per year. Going forward, I will analyze the case for the mismeasurement hypothesis using \$3 trillion as the implied value of output “lost” because of the productivity slowdown. This measure is conservative in the sense that it leaves less total lost output for the hypothesis to explain than would applying the BLS measured productivity slowdown to all of GDP. Based on 2015 US Census estimates of a US population of 321 million living in 125 million households, this works out to output that is lower because of the productivity slowdown by \$9,300 per capita and \$24,100 per household.

Thus, to explain the entire productivity slowdown as a figment of measurement problems implies that every person in the United States in 2015 enjoyed an average additional surplus of \$9,300 that did not exist in 2004.

It is important to recognize that the question is not whether the average consumer surplus in 2015 is \$9,300 per capita. GDP does not measure, nor ever has measured, consumer surplus. Nominal GDP values output at its market price; consumer surplus is the extent to which willingness to pay is above the market price. There surely was consumer surplus in both 2004 and 2015, and it was probably substantial in both years. The question instead is whether

⁵²I thank Robert Gordon for sharing the hours data.

it is plausible that technological growth between 2004 and 2015—and in particular the advent and diffusion of digitally oriented technologies like smartphones, downloadable media, and social networks that have been the most cited examples—created \$9,300 per person in incremental and unmeasured value above and beyond any consumer surplus that already existed in goods and services present in 2004 and was brought forward to 2015.

The Extent of the Productivity Slowdown Is Not Related to Digital Technology Intensity

Several studies have noted recent productivity slowdowns in economically advanced countries (for example, Mas and Stehrer 2012; Connolly and Gustafsson 2013; Pessoa and Van Reenen 2014; Goodridge, Haskel, and Wallis 2015). As in the US economy, these slowdowns began before the 2008–09 financial crisis and recession (Clette, Fernald, and Mojón 2015).

Given the relatively technology-heavy profile of US production (and citation of digital technologies produced by US-based multinationals as prime examples of the sources of mismeasurement), one might argue that the fact that a productivity slowdown has occurred across a number of economies makes a measurement-based explanation for the slowdown less likely. Still, similar measurement problems could have arisen in multiple advanced economies. I test if there is any systematic relationship between the extent of a slowdown in a country and the importance of information and communications technology (ICT), whether on the production or consumption side, to that country's economy. The logic of this test is, if information and communication technologies have caused measured productivity to underestimate true productivity, the mismeasurement hypothesis would imply that the measured slowdown in productivity growth should be larger in countries with greater “ICT intensity.”

I conduct this test using OECD labor productivity growth data, which contains yearly percentage changes in real GDP per worker-hour. Growth rates are reported for about three dozen countries in 2015—the latest year for which data are available—but only 30 have data going back to 1995 as needed to directly compare to the US slowdown. I combine this productivity growth data with two measures, also from the OECD, of the intensity of an economy in information and communications technology. The consumption-side measure is the fraction of a country's households with broadband internet access. My data are taken from 2007, the year in which this data was most widely available, and cover 28 countries, 25 of which overlap with those for which I can compute the change in average annual productivity growth between

1995–2004 and 2005–2015.⁵³ Obviously broadband access has increased since this time, but here I am interested in the much more stable cross-sectional variation. The production-side intensity metric is the share of the country's added value accounted for by industries related to information and communications technology. These data are only available for 2011. They span 28 countries, 24 of which overlap with my productivity slowdown sample.

The ubiquity of the productivity slowdown is readily apparent in the data. Labor productivity growth decelerated between 1995–2004 and 2005–2015 in 29 of the 30 countries in the sample (Spain is the only exception). Labor productivity growth across the sample's countries fell on average by 1.2 percentage points per year between the periods, from 2.3% during 1995–2004 to 1.1% over 2005–2015. There was substantial variation in the magnitude of the slowdown, with a standard deviation of 0.9% per year across countries. While the crisis years of 2008–2009 saw unusually weak productivity growth—these were the only two years with negative average productivity growth across the sample—the slowdown does not merely reflect the crisis years. Calculating later-period average productivity growth excluding 2008–2009 still reveals slowdowns in measured productivity growth in 28 of 30 countries (excepting Spain and Israel), with an average drop of 0.9 percentage points per year (a decline in annual rates from 2.3% to 1.4%). Similarly, computing the prior period average productivity growth using only 1996–2004 data in order to allow for an expanded sample gives the same results: productivity growth slows between the periods in 35 of 36 countries (Spain is again the exception).

To consider the covariance between the size of a country's slowdown and its information and communications technology (ICT) intensity, **Figure 1**, Panel A plots each country's change in average annual labor productivity growth between 1995–2004 and 2005–2015 against the share of the country's households that have broadband access. There is no obvious relationship to the eye, and this is confirmed statistically. Regressing the change in labor productivity growth on broadband penetration yields a coefficient on broadband of -0.0003 (s.e. = 0.009). The point estimate implies that a one standard deviation difference in broadband penetration is associated with less than a one-hundredth of a standard deviation difference in the magnitude of the slowdown.

On the production side, Figure 1, Panel B, plots the change in average annual labor productivity versus the share of a country's added value due to

⁵³Two countries, Iceland and Turkey, did not have 2015 data available, so I instead use 2005–2014 as the later period. I also use 2005–2014 for Ireland because reported labor productivity growth in 2015 was 22.5%, an astonishing number and one that is likely due to tax-driven corporate inversions (for example, Doyle 2016). That said, the results are not sensitive to these substitutions.

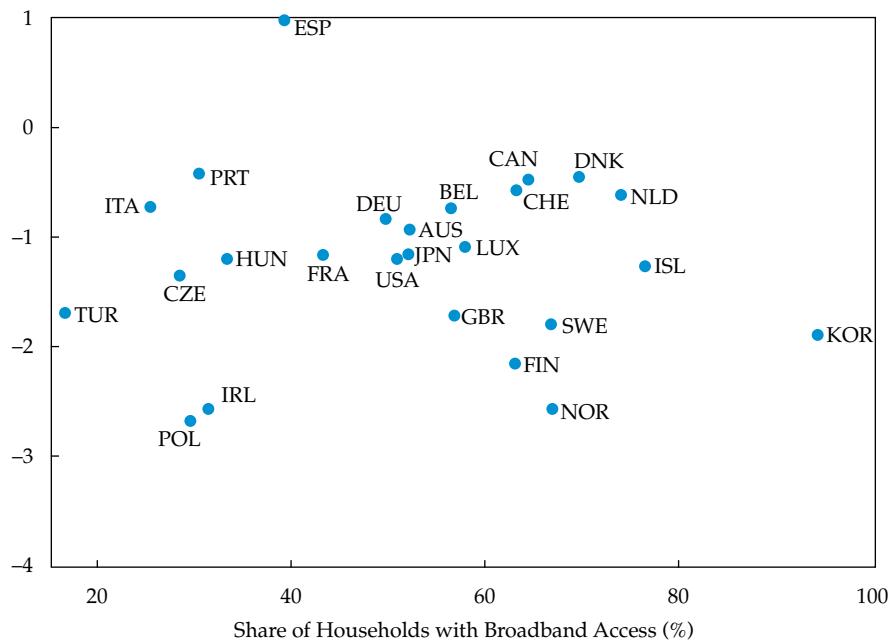
its ICT industries. Here the visual is less obvious, but as with the previous panel, a regression yields a statistically insignificant relationship. The coefficient on intensity of production in information and communications technology is -0.123 (s.e. = 0.101). To the extent any relationship exists, it is due completely to the outlier Ireland, which has a value-added share in information and communications technology of 11.9%, double the sample average. Removing Ireland from the sample yields a statistically insignificant coefficient of -0.054 (s.e. = 0.133). This point estimate correlates a one standard deviation difference in share of value added from information and communications technology to one-eleventh of a standard deviation change in the magnitude of the productivity slowdown.

Similar results obtain both qualitatively and quantitatively if I instead measure the productivity slowdown using later-period growth rates that exclude 2008–2009 or the larger sample with 1996–2004 as the early period.

Figure 1. Change in Labor Productivity Growth versus Information and Communication Technology (ICT) Intensity

A. Labor Productivity Growth Change between 1995–2004 and 2005–2015 versus Share of Households with Broadband Access (N = 25 OECD countries)

Percentage Point Change in Labor Productivity Growth

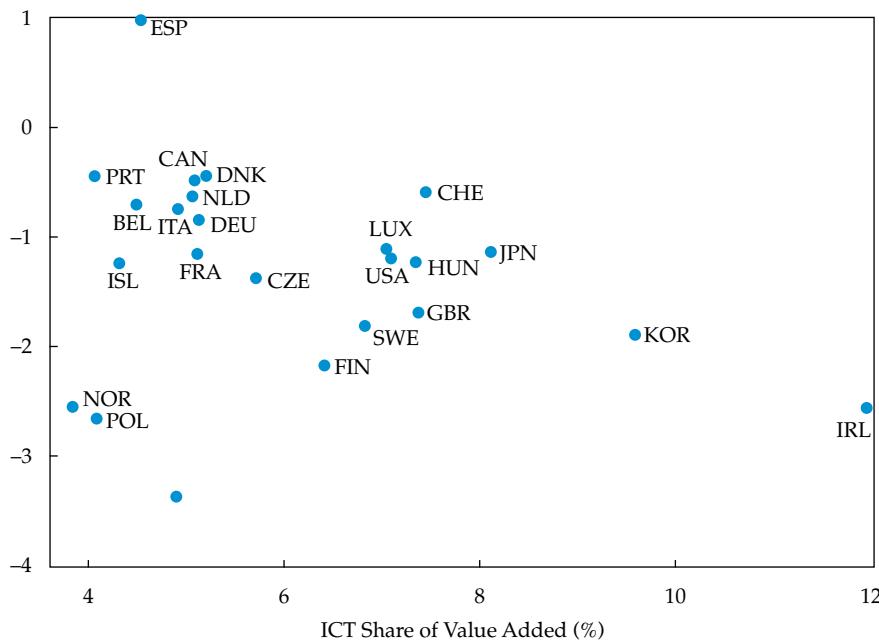


(Continued)

Figure 1. Change in Labor Productivity Growth versus Information and Communication Technology (ICT) Intensity (Continued)

B. Labor Productivity Growth Change between 1995–2004 and 2005–2015 versus ICT's Share of Value Added (N = 24 OECD countries)

Percentage Point Change in Labor Productivity Growth



Source: Data for both figures are from OECD. See text for details.

This is not surprising given that the correlations between the three productivity slowdown measures are all above 0.9.

Overall, the size of the productivity slowdown in a country does not seem to be systematically related to measures of the intensity of consumption or production of information and communications technology in that country. These results echo and complement the findings of Cardarelli and Lusinyan (2015), who show that differences in the slowdown in total factor productivity growth across US states are uncorrelated with measures of state-level intensity of information and communication technologies, both as inputs and outputs in production.

Estimates of Surplus from Internet-Linked Technologies

Several researchers have attempted to measure the consumer surplus of newer technologies like those discussed in the context of the mismeasurement

hypothesis. While not always explicitly motivated by the post-2004 measured productivity slowdown (some of these studies predated the recognition of the productivity slowdown among scholars), these analyses were impelled by a similar notion: Certain newer technologies, those tied to internet access in particular, may have an exceptionally high ratio of consumer surplus/observed expenditure. Several studies that seek estimates of these values, which I update here, offer insight into the potential for such technologies to explain the productivity slowdown.

Greenstein and McDevitt (2009) estimate the consumer surplus created by broadband access. They choose broadband because as an access channel, its price at least partially embodies the surplus created by otherwise unpriced technologies (for example, internet search, some downloadable media, social networking sites, and others). As Greenstein (2013, p.11) notes, “Looking at broadband demand, which does have a price, helped capture the demand for all the gains a user would get from using a faster form of Internet access.” They estimate that the new consumer surplus created by households that switched from the earlier technology (dial-up) was between 31% and 47% of broadband’s incremental revenue over dial-up. At the end of their analysis sample in 2006, this consumer surplus totaled \$4.8 billion to \$6.7 billion. In 2015, total US broadband revenues are estimated to be \$55 billion (see The Statistics Portal, <http://www.statista.com/statistics/280435/fixed-broadband-access-revenues-in-the-united-states>).

Supposing broadband’s overall ratio of consumer surplus/revenues is the same in 2015 as Greenstein and McDevitt (2009) estimated, this implies that the consumer surplus of broadband was \$17 billion to \$26 billion in 2015. Some of this value is likely priced into GDP indirectly through broadband’s use by producers as an intermediate input and as such should not be considered part of the missing output due to the productivity slowdown. But even absent any such adjustment, this surplus is two orders of magnitude smaller than the \$3 trillion of missing output.

Dutz, Orszag, and Willig (2009) apply demand estimation techniques to household data on internet service take-up and prices. They estimate a consumer surplus from broadband (again relative to dial-up) on the order of \$32 billion per year in 2008. To scale up this value for the growth in broadband since then, I use the fact that their estimates implied the same consumer surplus was \$20 billion in 2005. Assuming this robust 60% growth over three years (a compounded annual growth rate of 17%) held until 2015, consumer surplus in 2015 would be \$96 billion. While this is notably larger than the Greenstein and McDevitt (2009) valuation, it is still only 3.2% of \$3 trillion.

In another attempt to measure broadband's consumer surplus, Rosston, Savage, and Waldman (2010) use a different methodology and dataset. Their estimate is \$33.2 billion in 2010. I bring this forward to 2015 using their assessment that this surplus had doubled or perhaps even tripled between 2003 and 2010, which implies a compound annual growth rate between 10.4% and 17.0% (which, as it happens, is on the order of the growth rate in Dutz, Orszag, and Willig 2009). This extrapolation implies consumer surplus was in the range of \$54 billion to \$73 billion in 2015. Once again, this is minuscule compared to the lost output.

Nevo, Turner, and Williams (2015) use household-level data on broadband purchases to estimate a dynamic model of broadband demand. They find an average consumer surplus among households in their data between \$85 and \$112 per month (\$1,020 to \$1,344 per year) in 2012. Applying this to the 80% of US households that had broadband access in 2015, this totals at most \$132 billion—larger than the estimates above, but again less than 5% of the \$3 trillion in missing GDP.⁵⁴

Goolsbee and Klenow (2006) take a different approach. They use the time people spend online as an indicator of “full expenditure” on internet-based technologies. In their methodology, consumption of a good generally involves expenditure of both income and time. Therefore, even if financial expenditures on a good are relatively small, the good can deliver substantial welfare if people spend a lot of time consuming it. They argue this is a realistic possibility for the internet, which in their data (for 2005) has a time expenditure share 30 times greater than its income expenditure share. Applying their theoretical framework to data, they find that the consumer surplus of internet access could be as large as 3% of full income (the sum of actual income and the value of leisure time). This surplus would be \$3,000 annually for the median person in their dataset. Brynjolfsson and Oh (2012) extended this analysis with updated data. They pay particular attention to incremental gains from free internet services, valuing these at over \$100 billion (about \$320 per capita) annually.

To extend the Goolsbee and Klenow (2006) value-of-time analysis to the question of the mismeasurement hypothesis, I must first compute total income in 2015. Disposable personal income totaled \$13.52 trillion, about

⁵⁴They also use their estimates to infer the total surplus (revenues plus consumer surplus) of access to 1 Gb/s networks, which is currently unavailable in most locations. This extrapolation implies a total surplus of \$3,350 per year. Some of this would surely be captured as revenues of downstream firms and thus measured in GDP. A conservative price for this service would be \$900 per year, so consumer surplus per household would be around \$2,450. Even if service were obtained by every household in the country that has broadband, this adds up to \$241 billion of consumer surplus, which is 8% of \$3 trillion.

\$42,100 per capita, in 2015. For the value of leisure time, I start with the fact that according to the American Time Use Survey (ATUS), the average person in 2014 spent 10.8 hours a day on non-work-related, non-personal-care activities. (Personal care includes sleep, so sleep is not included in the 10.8 hours.) I make the (very) generous assumptions that all these 10.8 hours are leisure time and that people value them at the average after-tax wage of \$22.08, regardless of employment status and whether the hours are inframarginal or marginal. This value of time is based on the estimate by the Bureau of Labor Statistics that average pretax hourly earnings for all nonfarm private business employees were \$25.25 over the final quarter of 2015. To impute an after-tax wage, I multiply this value by the ratio of that quarter's disposable personal income (\$13.52 trillion) to total pretax personal income (\$15.46 trillion), reflecting an average tax rate of 12.5%. This yields a total annual value of leisure time of about \$87,000 per person. Adding this to personal income gives a total income equal to \$129,100 per capita.

Applying the Goolsbee and Klenow (2006) top-end estimate that it is 3% of total income, I end up with a measure of the consumer surplus from the internet in 2015 of around \$3,900 per capita.⁵⁵ Assuming this surplus accrues mainly to the 80% of people with broadband access in their household, the aggregate benefit is \$995 billion. Going through the same set of computations with 2004 data (when broadband penetration was about 12% according to OECD data) and subtracting the result so as to estimate incremental gains from broadband-based technologies yields a post-2004 incremental surplus from broadband of \$863 billion.⁵⁶

⁵⁵As noted in the text, the 3% value is determined in part from Goolsbee and Klenow's (2006) time use data. It is plausible that the ratio of the internet's time expenditure share to its income expenditure share could have risen in the intervening decade, thereby raising this number. However, comparable contemporaneous data necessary to check this is difficult to find. The ATUS does not offer a separate item for online activity save for an email category that accounts for a tiny share of time. Many commercially available data products do not separate online leisure from online work time (the latter being an input into production rather than a final output) and allow multitasking, so a day can be filled with more than 24 hours of activity. In absence of specific guidance, I keep the original 3% value here.

⁵⁶The specific figures for 2004 are \$9 trillion of nominal disposable income (\$30,700 per capita given a population of 293 million), 11 hours of leisure time per day, and \$18.19 per hour after-tax nominal hourly earnings (based on Bureau of Labor Statistics earnings data for 2006, the start of the all-worker-compensation series). This implies a total nominal income of \$103,800 per capita. Applying the 2004–15 GDP deflator ratio of 1.21 and multiplying by the Goolsbee–Klenow estimate of 3% yields a benefit of \$3,800 per capita in 2015 dollars. This is very close to the 2015 figure, so almost all incremental surplus from broadband by this calculation comes from diffusion of broadband to a larger population. This increase in population with broadband is $(0.8 \times 321 \text{ million}) - (0.12 \times 293 \text{ million}) = 222 \text{ million}$.

The Goolsbee–Klenow time-based estimate is by far the highest valuation of the internet in the literature, essentially an order of magnitude larger than the other estimates. Time-of-use valuation approaches can produce large numbers; there are always 24 hours in a day to allocate and value, and it is hard to estimate the monetary value of a minute. Indeed, one could have used a similar logic to argue that productivity numbers in the 1950s and 1960s—the height of the post–World War II productivity acceleration—were missing the allegedly massive social gains of families’ fast-increasing TV viewing. I stick with common practice and apply a (generous) wage-related valuation here, but in principle the wage applies only to the unit of time on the margin of work. Inframarginal leisure time should be valued by the incremental surplus relative to the next-best use of that time: for example, the extra amount someone is willing to pay to be online as opposed to, say, watch television. This increment could be much smaller than the person’s wage, and the increment and wage may be uncorrelated across people, making the \$863 billion figure a large overstatement. Even given these measurement issues, the implied valuation from the time-of-use approach is still less than one-third the \$3 trillion of lost income from the productivity slowdown.

Most of the technologies cited by proponents of the mismeasurement hypothesis require internet access of some sort, so these estimates of the surplus delivered by that gateway should embody the surplus of the technologies that are not priced on the margin. It is possible that some post-2004 technologies that deliver a high ratio of consumer surplus to revenue do not require internet access. The numbers above indicate, however, that to explain the bulk of the productivity slowdown in quantitative terms, these products would need to deliver surplus that is both somehow not priced either directly or through complementary goods and services and that is as large as or larger than the biggest estimates of the surplus of internet-linked products.

What If the “Missing” Output Were Measured?

Yet another calculation of the quantitative plausibility of the mismeasurement hypothesis relates the \$3 trillion of missing GDP to the added value of the specific products associated with post-2004 technologies. I take an expansive view of which products include such technologies, in an attempt to construct something of an upper bound of the lost output that can be explained by the hypothesis.

The first step in this calculation is to select the set of technologies that would be most implicated in the mismeasurement if GDP mismeasurement results from the migration of value from output to consumer surplus since 2004. I include the following sectors in this group: computer and

electronic products manufacturing (NAICS 334), the entire information sector (NAICS 51), and computer systems design and related services (NAICS 5415). The first and last are self-explanatory. The information sector includes the following four subindustries: publishing (including software), except internet; motion picture and sound recording; broadcasting and telecommunications; and data processing, internet publishing, and other information services. Both internet service providers and mobile telephony carriers are in this sector (in particular, NAICS 517, telecommunications).

These industries comprise the segments of the economy most likely to produce the technologies that are the focus of claims of the mismeasurement hypothesis. They also doubtlessly contain some activity that has not seen considerable technological expansion over the past decade (or even the past couple of decades, for that matter). As will be clear, this overexpansive definition of the output tied to the mismeasurement hypothesis is conservative in the sense that it will tend to overestimate the missing output of these industries for which technological developments in these industries might account.

The added value of these industries in 2015 was as follows: computer/electronics manufacturing, \$278 billion; information, \$840 billion; computer systems design and services, \$266 billion. This totals \$1,384 billion.

At the precipice of the productivity slowdown in 2004, the nominal added value of the sectors was \$945 billion (\$202 billion in computer/electronics manufacturing, \$621 billion in information, and \$123 billion in computer systems design and services). Applying the Bureau of Economic Analysis value-added price indices of the three sectors yields 2004 added value expressed in 2015 dollars: \$813 billion.⁵⁷

These industries therefore saw measured real value-added growth between 2004 and 2015 of about \$571 billion (that is, \$1,384 billion – \$813 billion). If measurement problems in the products of these industries are to account for the lion's share of \$3 trillion in missing GDP, the incremental consumer

⁵⁷This method divides the industries' summed nominal value added in 2004 by a Tornqvist price index I constructed for the combined industries. This index is equal to the average-share-weighted sum of the log changes in each of the three components' price indexes from 2004 to 2015. Note that all three industries saw drops in their value-added price indices over the period, which is why the figure in 2015 dollars is smaller than the 2004 figure. An alternative approach of deflating each industry's 2004 nominal added value by the industry-specific deflator and summing the result implies 2004 real value added in 2015 dollars of \$829 billion. The difference in the methods mostly reflects the effect of the 36% decline in the computer equipment manufacturing price index during the period. Note that using this latter figure for 2004 added value in the calculations below would make the "missing" output of the mismeasurement hypothesis even larger in terms of the industries' measured incremental added value.

surplus these industries would have created would need to be over six times their measured incremental added value. Or to put this another way, if the incremental consumer surplus implied by the mismeasurement hypothesis would in fact have been captured as measured added value (and therefore the productivity slowdown observed in the data never materialized), the real value added of the industries would actually have increased by 440% ($(\$1.384 \text{ trillion} + \$3 \text{ trillion})/\$813 \text{ billion}$), over six times the 70% growth ($\$1.384 \text{ trillion}/\813 billion) that was actually observed in the data. This implies an enormous amount of mismeasurement. Even to account for just one-third of the missing output, by far the largest estimate of surplus from internet-related products discussed in the prior section, the industries’ “correct” added value would have had to have grown by 190% from 2004–2015, almost triple the measured growth.

Looking at the dual to this calculation—that is, not how much larger the “real” output would need to be, but how much larger the price deflator would need to be—is also instructive. The (Tornqvist) value-added price index for this bundle of industries fell 14% over 2004–2015, a compound annual growth rate of -1.4%. If real GDP growth has been misstated because deflators have improperly accounted for quality changes in these products, the true deflator would be that which raises measured real value-added growth by the extra \$3 trillion. This deflator would have a compound annual growth rate of -9.9%, sustained over 11 years—seven times the magnitude of the official deflator. Prices would have fallen not by 14% since the productivity slowdown began but by 68% instead. Some of the outputs of these industries are intermediate inputs used to make other products. Therefore, they do not directly deliver surplus to final demanders. It is possible that some of the gains from the new technologies might arise as (again mismeasured) productivity gains in the production of goods for which they are used as inputs. For example, in the 2015 input-output tables for the national income and product accounts, 83% of computer equipment manufacturing output was used as an intermediate in the production of another commodity. The corresponding values for information and computer services are 46% and 42%, respectively. The total “multiplier” effect of technological progress through input use is captured by the industry’s ratio of gross output (revenues) to its added value (Domar 1961; Hulten 1978). Incremental revenues capture the gains associated with not just the industry’s products per se but also any embodied productivity gains obtained through their use as inputs. To gauge the potential influence of this usage, I repeat the calculations above using revenues—that is, gross output—in place of added value.

The nominal gross output of the three sectors in 2015 was \$2.29 trillion (\$387 billion in computer and electronics manufacturing, \$1,550 billion in

information, and \$353 billion in computer systems design and services). The corresponding values in 2004 were \$1.67 trillion (\$392 billion, \$1,080 billion, and \$195 billion). Again applying the Bureau of Economic Analysis price deflators (this time for gross output) to express these values in 2015 dollars yields a real gross output of \$1.61 trillion.

Incremental real gross output (that is, real revenue) for this set of industries was therefore about \$680 billion. A full accounting for the mismeasurement hypothesis would imply an increment to consumer surplus that is five times as large as this. Had such a surplus been captured in revenue figures, the industries' real revenues would have more than tripled over 2004–2015, rather than risen 42% as observed in the data. The dual calculation implies a mismeasurement-corrected deflator with a compound average growth rate of -7.3% over 2004–2015 instead of the official gross output price index compound average growth rate of -0.3%, for a total price decline of 57% rather than 3%.

These calculations reveal how severely one must believe the measured growth of these industries understates their true growth if measurement problems are to explain the overall productivity slowdown for the entire US economy. What was measured and what would have actually had to happen would be multiples apart.

A final set of calculations reinforces this point. If the data miss industry output growth, they of course also miss productivity growth. In this case, it would need to be a lot of missing productivity. These industries, combined, saw their total employment rise 3.2% over 2004–2015 (from 5.58 million to 5.76 million, about 0.3% annually). Assuming they actually produced all of the output lost to the productivity slowdown, real value added per worker, properly measured, would have risen by 415% over those 11 years, an astounding rate of productivity growth. For example, it is notably larger than the 83% productivity growth seen in durable goods manufacturing during the productivity acceleration of 1995 to 2004, when durables had the fastest labor productivity growth of any major sector and they were a primary driver of the acceleration (Oliner, Sichel, and Stiroh 2007).

Perhaps these numbers are not that surprising when one considers that these digital-technology industries accounted for only 7.7% of GDP in 2004. A full accounting of the productivity slowdown by the mismeasurement hypothesis requires this modest share of economic activity to account for lost incremental output that in 2015 is about 17% of GDP—over twice the 2004 size of the entire sector.

One should be mindful that it is possible that unmeasured incremental gains are being made in industries outside these. For example, more intensive

use of information technologies has been a recent focus of attention (including public policy efforts) in the sizable health care sector. Yet evidence on the productivity benefits of specific technologies in the sector has been mixed (for example, Agha 2014; Bhargava and Mishra 2014). There does not appear to be a clear case for large missing gains in the sector. Moreover, further balancing this out is the fact that as discussed above, the digital-product-focused industries here are defined expansively. It is unlikely that every segment in this grouping (as one example, radio broadcasting) experienced similarly rapid technological progress.

National Income versus National Product

In national income accounting, it is an identity that gross domestic product (GDP) is equal to gross domestic income (GDI)—the sum of employee compensation, net operating surplus, net taxes on production and imports, and consumption of fixed capital (that is, depreciation). However, GDP and GDI are never equal in practice, because different data are used to construct each—expenditure data on the one hand and income information on the other.

In recent years, the gap between GDI and GDP—the so-called “statistical discrepancy”—has widened, with GDI on average outpacing GDP.

Table 2 shows GDI, GDP, and the gap between them in annual data for 1995–2015.⁵⁸ Over 2005–2015, a cumulative gap of \$903 billion (nominal) grew between GDI and GDP. This is an average gap of about 0.5% of GDP per year, though not every single year saw domestic income exceed domestic product. One could argue that this gap reflects workers being paid to make products (whose labor earnings are included in GDI) that are being given away for free or at highly discounted prices relative to their value (reducing measured expenditures on these products and therefore GDP in turn). This would be an indicator of the forces surmised by the mismeasurement hypothesis.

A closer examination of the data, however, strongly suggests that the GDI–GDP gap is not a sign of the mismeasurement hypothesis.

First, the gap started opening before the productivity slowdown. GDI was larger than GDP in each of the seven years running from 1998 to 2004, all of which were a time of fast productivity growth. The average annual gap was 0.6% of GDP, even larger than in the slowdown period.

⁵⁸The US Bureau of Economic Analysis defines the statistical discrepancy as GDP minus GDI, so a negative reported value implies that GDI is larger than GDP. I am focusing on the extent to which GDI is greater than GDP, so I am discussing the behavior of the negative of the statistical discrepancy.

Table 2. Gross Domestic Income versus Gross Domestic Product

Year	GDI (\$ billions)	GDP (\$ billions)	GDI–GDP Gap (\$ billions)	Percentage of GDI Going To			
				Labor Income	Net Operating Surplus	Net Taxes	Depreciation
1995	7,573.5	7,664.1	–90.6	55.5	22.7	6.9	14.8
1996	8,043.6	8,100.2	–56.6	55.0	23.6	6.8	14.6
1997	8,596.2	8,608.5	–12.3	54.9	24.0	6.7	14.4
1998	9,149.3	9,089.2	60.1	55.5	23.5	6.6	14.3
1999	9,698.1	9,660.6	37.5	55.9	23.2	6.5	14.4
2000	10,384.3	10,284.8	99.5	56.5	22.6	6.4	14.6
2001	10,736.8	10,621.8	115	56.4	22.4	6.2	14.9
2002	11,050.3	10,977.5	72.8	55.7	22.8	6.5	15.0
2003	11,524.3	11,510.7	13.6	55.3	23.1	6.6	15.0
2004	12,283.5	12,274.9	8.6	54.9	23.5	6.7	14.9
2005	13,129.2	13,093.7	35.5	54.1	24.2	6.7	15.1
2006	14,073.2	13,855.9	217.3	53.4	24.7	6.7	15.2
2007	14,460.1	14,477.6	–17.5	54.7	22.9	6.8	15.7
2008	14,619.2	14,718.6	–99.4	55.3	21.7	6.8	16.2
2009	14,343.4	14,418.7	–75.3	54.4	22.4	6.7	16.5
2010	14,915.2	14,964.4	–49.2	53.4	23.9	6.7	16.0
2011	15,556.3	15,517.9	38.4	53.2	24.3	6.7	15.8
2012	16,358.5	16,155.3	203.2	52.7	25.3	6.6	15.5
2013	16,829.5	16,691.5	138	52.6	25.2	6.6	15.6
2014	17,651.1	17,393.1	258	52.5	25.4	6.5	15.6
2015	18,290.3	18,036.6	253.7	53.1	25.0	6.5	15.5

Note: Data are from the US Bureau of Economic Analysis, national income accounts Table 1.10.

Second, a closer look at the composition of national income reveals patterns inconsistent with the “workers paid for making free (or nearly free) products” story.

The four right-most columns in Table 2 follow the evolution of the shares of GDI paid to each of the four major income categories that comprise it. Between 2004 and 2015, employee compensation’s share of GDI fell by 1.8 percentage points, while net operating surplus grew by 1.5 percentage points. The net taxes share fell by 0.2 percentage points and depreciation rose by 0.6 percentage points. Thus, the GDI gains over the period were tied to

payments to capital that came at the expense of labor income.⁵⁹ Nor is this link between GDI and capital income only manifested in long differences; the correlation in annual data from 1995 to 2015 between the GDI–GDP percentage gap and labor's share is -0.35 , while it is 0.58 for net operating surplus.

Growth in domestic income measures relative to measured domestic product therefore seems to reflect increases in capital income rather than labor income. “Abnormally” high measured income relative to measured expenditures is positively related to growth in businesses’ profitability and negatively related to payments to employees. This is inconsistent with—and indeed implies the opposite of—the “pay people to build free goods” story.

Conclusion

What I have termed the “mismeasurement hypothesis” argues that true productivity growth has not slowed (or has slowed considerably less than measured) since 2004, but recent gains have not been reflected in productivity statistics, either because new goods’ total surplus has shifted from (measured) revenues to (unmeasured) consumer surplus, or because price indices are overstated. My evaluation focuses on four pieces of evidence that pose challenges for mismeasurement-based explanations for the productivity slowdown that the US economy has been experiencing since 2004. Two patterns—the size of the slowdown across countries is uncorrelated with the information and communications technology intensities of those countries’ economies, and the GDI–GDP gap began opening before the slowdown and in any case reflects capital income growth—are flatly inconsistent with the implications of the mismeasurement hypothesis. Two others—the modest size of the existing literature’s estimates of surplus from internet-linked products and the large implied missing growth rates of digital-technology industries that the mismeasurement hypothesis would entail—show the quantitative hurdles the hypothesis must clear to account for a substantial share of what is an

⁵⁹These income share changes are a reflection of the trends that other researchers have been exploring in other contexts (for example, Elsby, Hobijn, and Sahin 2013; Karabarbounis and Neiman 2014). An alternative decomposition of income yields the same implications as those described here. This alternative divides national income (gross domestic income adjusted for international transfers minus depreciation) into employee compensation, proprietor’s income, capital income (the sum of rental income, corporate profits, and net interest), and a residual category that is the sum of net taxes on production and imports plus business transfer payments plus the surplus of government enterprises. As with the results above, labor’s share fell as capital’s share rose over 2004–2015. Employee compensation’s share of national income fell by 2.1 percentage points while capital income grew by 2.5 percentage points. (Proprietors’ income share fell by 0.3 percentage points, and the share of taxes fell by 0.1 percentage point over the period.)

enormous amount of measured output lost to the slowdown (around \$9,300 per person per year).

These results do not definitively rule out the possibility that productivity measurement problems may have developed over the past decade for specific products or product classes. However, the combined weight of the patterns presented here makes clear that the intuitive and plausible empirical case for the mismeasurement hypothesis faces a higher bar in the data, at least in terms of its ability to account for a substantial portion of the measured output lost to the productivity slowdown.

In addition to the quantitative analyses above, several qualitative points further bolster the case for skepticism about the mismeasurement hypothesis.

As briefly mentioned above, concerns about GDP mismeasurement preceded the recent slowdown, particularly regarding GDP's disconnect with social welfare. Perhaps, the argument goes, even if true productivity growth has slowed, it need not be the case that welfare growth has. I agree that GDP does not measure social welfare; it was not designed to do so. But the GDP-welfare disconnect is not a recent phenomenon. The mere fact that GDP is an imperfect measure of welfare is insufficient as evidence for the measurement hypothesis; instead, to support the hypothesis one must argue that a break in the GDP-welfare disconnect somehow developed around 2004. None of the evidence presented above indicates this. In fact, the estimates of the benefits of internet-linked technologies are measures of consumer surplus, which by definition are not in GDP. In other words, even if all that surplus (recall the largest estimate is \$863 billion) were somehow captured in GDP—which is not typically the case—it would still fall considerably short of making up for the GDP lost because of the productivity slowdown.

A second point is that my four analyses took as given the possibility that as the mismeasurement hypothesis asserts, many new goods post-slowdown are missed in GDP because of low or zero prices. However, it is not clear at all that this baseline assertion is correct. To enjoy all these free goods—Facebook, the camera on your phone, Google searches, and so on—one must purchase complementary goods: a smartphone, an iPad, broadband access, mobile telephony, and so on. If companies that sell those complements know what they are doing, they ought to be pricing the value of those “free goods” into the price of the complementary products. Their value ought to be captured in the product accounts through the prices of the complementary products that are required to consume them. As an example, at least one of these complementary goods sellers, Apple, has been famously profitable during the slowdown.

Finally, in parallel with this study, other researchers have been conducting independent work that also looked at the mismeasurement hypothesis.

Their approaches used different methods and data than mine, yet they came to the same conclusion. I mentioned earlier the work by Cardarelli and Lusinyan (2015), which shows that the differing rates of productivity slowdown across US states are not related to variations in the intensity of information and communications technology production across states. Nakamura and Soloveichik (2015) estimate the value of advertising-supported internet consumer entertainment and information. They apply the existing procedures for valuing advertising-supported media content in GDP and find that accounting for free-to-consumers content on the internet raises GDP growth by less than 0.02% per year. Byrne, Fernald, and Reinsdorf (2016) offer two main arguments. First, they readily admit that information technology hardware is mismeasured since 2004, but they argue that the mismeasurement was even larger in the 1995–2004 period. Moreover, more of the information technology hardware was produced in the United States in the 1995–2004 period. Taken together, these adjustments imply that the slowdown in labor productivity since 2005 looks worse, not better. The second main point is that consumers are using many information and communications technologies to produce service for their nonmarket time, which means that consumers benefit, but gains in nonmarket production (which in any event are small) do not suggest that market-sector productivity is understated.

If the theory that new products caused the productivity slowdown is to be resurrected, it may well need to take on a different form. For example, one very speculative mechanism that would tie a true productivity slowdown to people spending a large share of their time on zero-to-low-marginal-price activities would be if workers substituted work effort for technology consumption—for example, spending time while they are at work on social networking sites. This pattern would heighten consumer surplus in a way largely unmeasured by standard statistics while at the same time reducing output per hour—that is, measured labor productivity. Of course, to explain a slowdown in annual labor productivity growth, this substitution would need to be occurring in ever-greater magnitudes over time.

The empirical burdens facing the mismeasurement hypothesis are heavy, and more likely than not, much if not most of the productivity slowdown since 2005 is real. Whether that slowdown will end anytime soon remains an open question.

I thank Erik Brynjolfsson, Dave Byrne, Austan Goolsbee, Bob Gordon, Jan Hatzius, Pete Klenow, Rachel Soloveichik, Hal Varian, and the JEP editors for comments. I have no financial interests relevant to this study.

References

Agha, Leila. 2014. “The Effects of Health Information Technology on the Costs and Quality of Medical Care.” *Journal of Health Economics* 34 (March): 19–30.

Alloway, Tracy. 2015. “Goldman: How ‘Grand Theft Auto’ Explains One of the Biggest Mysteries of the U.S. Economy.” *Bloomberg* (26 May). <http://www.bloomberg.com/news/articles/2015-05-26/goldman-how-grand-theft-auto-explains-one-of-the-biggest-mysteries-of-the-u-s-economy>.

Baily, Martin Neil, James Manyika, and Shalabh Gupta. 2013. “U.S. Productivity Growth: An Optimistic Perspective.” *International Productivity Monitor* 25 (Spring): 3–12.

Becker, Gary S. 1965. “A Theory of the Allocation of Time.” *Economic Journal* 75 (299): 493–517.

Bhargava, Hemant K., and Abhay Nath Mishra. 2014. “Electronic Medical Records and Physician Productivity: Evidence from Panel Data Analysis.” *Management Science* 60 (10): 2543–62.

Brynjolfsson, Erik, and Andrew McAfee. 2011. *Race Against the Machine: How the Digital Revolution Is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*. Lexington, MA: Digital Frontier Press.

———. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W.W. Norton & Company.

Brynjolfsson, Erik, and JooHee Oh. 2012. “The Attention Economy: Measuring the Value of Free Digital Services on the Internet.” *Proceedings of the International Conference on Information Systems*, Orlando. <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1045&context=icis2012>.

Byrne, David M., John G. Fernald, and Marshall B. Reinsdorf. 2016. “Does the United States Have a Productivity Slowdown or a Measurement Problem?” *Brookings Papers on Economic Activity* 2016 (Spring): 109–57. <http://www.jstor.org/stable/43869021>.

Byrne, David M., Stephen D. Oliner, and Daniel E. Sichel. 2013. “Is the Information Technology Revolution Over?” *International Productivity Monitor* 25 (Spring): 20–36. <https://ideas.repec.org/a/sls/ipmsls/v25y20133.html>.

———. 2015. “How Fast Are Semiconductor Prices Falling?” NBER Working Paper 21074.

Cardarelli, Roberto, and Lusine Lusinyan. 2015. “U.S. Total Factor Productivity Slowdown: Evidence from the U.S. States.” IMF Working Paper 15/116.

Cette, Gilbert, John Fernald, and Benoît Mojon. 2015. “The Pre-Global-Financial-Crisis Slowdown in Productivity.” Unpublished paper.

Connolly, Ellis, and Linus Gustafsson. 2013. “Australian Productivity Growth: Trends and Determinants.” *Australian Economic Review* 46 (4): 473–82.

Cowen, Tyler. 2011. *The Great Stagnation: How America Ate All the Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better*. New York: Dutton.

Davis, Steven J., and John Haltiwanger. 2014. “Labor Market Fluidity and Economic Performance.” NBER Working Paper 20479.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2014. “The Role of Entrepreneurship in US Job Creation and Economic Dynamism.” *Journal of Economic Perspectives* 28 (3): 3–24.

Diewert, W. Erwin, and Kevin J. Fox. 1999. “Can Measurement Error Explain the Productivity Paradox?” *Canadian Journal of Economics* 32 (2): 251–80.

Domar, Evsey D. 1961. “On the Measurement of Technological Change.” *Economic Journal* 71 (284): 709–29.

Doyle, Dara. 2016. “Ireland’s Economists Left Speechless by 26% Growth Figure.” *Bloomberg* (12 July).

Dutz, Mark, Jonathan Orszag, and Robert Willig. 2009. “The Substantial Consumer Benefits of Broadband Connectivity for U.S. Households.” Commissioned by the Internet Innovation Alliance. https://internetinnovation.org/files/special-reports/CONSUMER_BENEFITS_OF_BROADBAND.pdf.

Elsby, Michael W.L., Bart Hobijn, and Ayşegül Şahin. 2013. “The Decline of the U.S. Labor Share.” *Brookings Papers on Economic Activity* 2013 (2): 1–63.

Feldstein, Martin. 2015. “The U.S. Underestimates Growth.” *Wall Street Journal* (18 May).

Fernald, John G. 2014a. “Productivity and Potential Output Before, During, and After the Great Recession.” NBER Working Paper 20248.

———. 2014b. “A Quarterly, Utilization-Adjusted Series on Total Factor Productivity.” Working Paper 2012-19, Federal Reserve Bank of San Francisco.

Goodridge, Peter, Jonathan Haskel, and Gavin Wallis. 2015. "Accounting for the UK Productivity Puzzle: A Decomposition and Predictions." Imperial College London Discussion Paper 2015/02.

Goolsbee, Austan, and Peter J. Klenow. 2006. "Valuing Consumer Products by the Time Spent Using Them: An Application to the Internet." *American Economic Review* 96 (2): 108–13.

Gordon, Robert J. 2016. *The Rise and Fall of American Growth: The U.S. Standard of Living Since the Civil War*. Princeton, NJ: Princeton University Press.

Greenstein, Shane. 2013. "Measuring Consumer Surplus Online." *Economist* (March): 11.

Greenstein, Shane, and Ryan C. McDevitt. 2009. "The Broadband Bonus: Accounting for Broadband Internet's Impact on U.S. GDP." NBER Working Paper 14758.

Hatzius, Jan, and Kris Dawsey. 2015. "Doing the Sums on Productivity Paradox v2.0." *Goldman Sachs U.S. Economics Analyst* 15/30.

Hulten, Charles R. 1978. "Growth Accounting with Intermediate Inputs." *Review of Economic Studies* 45 (3): 511–18.

Karabarbounis, Loukas, and Brent Neiman. 2014. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129 (1): 61–103.

Klenow, Peter J. 2003. "Measuring Consumption Growth: The Impact of New and Better Products." *Federal Reserve Bank of Minneapolis Quarterly Review* 27 (2): 10–23.

Mas, Matilde, and Robert Stehrer, eds. 2012. *Industrial Productivity in Europe: Growth and Crisis*. Northhampton, MA: Edward Elgar.

Mokyr, Joel. 2014. "Secular Stagnation? Not in Your Life." In *Secular Stagnation: Facts, Causes and Cures*, edited by Coen Teulings and Richard Baldwin, chapter 6. CEPR Press eBook. <https://voxeu.org/article/secular-stagnation-not-your-life>.

Nakamura, Leonard, and Rachel Soloveichik. 2015. "Valuing 'Free' Media across Countries in GDP." Federal Reserve Bank of Philadelphia Working Paper 15-25.

Nevo, Aviv, John L. Turner, and Jonathan W. Williams. 2015. "Usage-Based Pricing and Demand for Residential Broadband." NBER Working Paper 21321.

Oliner, Stephen D., Daniel E. Sichel, and Kevin J. Stiroh. 2007. "Explaining a Productive Decade." *Brookings Papers on Economic Activity* 2007 (1): 81–152. https://www.brookings.edu/wp-content/uploads/2007/03/2007a_bpea_oliner.pdf.

Pessoa, João Paulo, and John Van Reenen. 2014. "The UK Productivity and Jobs Puzzle: Does the Answer Lie in Wage Flexibility?" *Economic Journal* 124 (576): 433–52.

Rosston, Gregory L., Scott J. Savage, and Donald M. Waldman. 2010. "Household Demand for Broadband Internet in 2010." *B.E. Journal of Economic Analysis & Policy: Advances* 10 (1): 1–45.

Smith, Noah. 2015. "The Internet's Hidden Wealth." *Bloomberg Opinion* (10 June). <https://www.bloomberg.com/opinion/articles/2015-06-10/wealth-created-by-the-internet-may-not-appear-in-gdp>.

Syverson, Chad. 2013. "Will History Repeat Itself? Comments on 'Is the Information Technology Revolution Over?'" *International Productivity Monitor* 25: 37–40.

Tarullo, Daniel K. 2014. "Longer-Term Challenges for the American Economy." Speech given to 23rd Annual Hyman P. Minsky Conference: Stabilizing Financial Systems for Growth and Full Employment, Washington, DC, 9 April.

US Congress. 1996. "Toward a More Accurate Measure of the Cost of Living." Final Report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index. (The Boskin Commission Report.) 104th Congress, 2nd sess., S.Prt. 104-072. <http://www.ssa.gov/history/reports/boskinrpt.html>.

Difficulties and Diagnoses

Is America Still a Startup Nation?

Dane Stangler

Chief Policy Officer, Startup Genome

Is America still a “startup nation”? That question, posed to me for the Restoring American Economic Dynamism conference, is loaded with assumptions. One is that the United States was (and is) a nation defined by entrepreneurial dynamism. Another is that this characteristic may not persist. In this chapter, I explore entrepreneurship trends in the United States, factors behind those trends, and what the future may hold. Over the past decade, there has simultaneously been a decline in economic dynamism—especially along several indicators of entrepreneurship—yet a proliferation of efforts devoted to helping entrepreneurs. While over the past decade there has been considerable angst over the “vanishing” of American entrepreneurship, the entrepreneurial future of the United States is bright.⁶⁰ In fact, there is good reason for thinking we’re on the cusp of an entrepreneurial boom.

New Business Creation, Productivity, and Job Creation

Entrepreneurship is central to the American story, and to the meaning of American identity. We pride ourselves on a rich heritage of self-made men and women—those who “made America”—and on the ability of anyone, from anywhere or any background, to start and grow a business.⁶¹ While Israel is celebrated today as the “startup nation,” the United States not only has been synonymous with entrepreneurial success but also was, politically, a startup nation in the 18th century.⁶² So the question of whether or not this continues to be true—and how true it might be—is important economically, politically, and socially.

Parts of this essay are adapted from a work-in-progress book, *Startup Fever*, with thanks to the Smith Richardson Foundation for their generous support.

⁶⁰Leigh Buchanan, “American Entrepreneurship Is Actually Vanishing. Here’s Why,” *Inc.*, May 2015, <https://www.inc.com/magazine/201505/leigh-buchanan/the-vanishing-startups-in-decline.html>.

⁶¹Harold Evans, *They Made America: From the Steam Engine to the Search Engine, Two Centuries of Innovators* (New York: Back Bay, 2004).

⁶²Saul Singer and Dan Senor, *Start-Up Nation: The Story of Israel’s Economic Miracle* (New York: Twelve, 2009).

Any discussion of entrepreneurship must begin with language: Many different phrases are thrown around on this topic, and we need to be precise in what, exactly, we are referring to. There are new businesses, young firms, startups, “age 0” companies, the self-employed, business owners, entrepreneurs, establishments, and so on. Here, following the research literature, the discussion will concern only employer firms (those that have employees), and will use “new business” to refer to those companies that are less than a year old. These are also referred to by researchers as “age 0” firms. Young firms are typically defined as those that are less than 5 or 10 years old. The discussion here departs from the research literature, however, in that I use “startup” to refer specifically to technology-based (“high-tech”) businesses that seek to establish scalable business models. This accords more closely with the popular notions of a “startup.”

In the aggregate, new and young businesses are major drivers of job creation and productivity in the US economy. In part, this is because of their sheer number. On average, over the last 40 years, nearly half a million new businesses have been created every year. In 2016, for example (the latest year for which data are available), 433,000 new businesses (age 0) were formed, employing 2.57 million people. That same year, existing firms of every other age created, on net, 352,000 new jobs. Many of those jobs initially created by new businesses will eventually disappear as firms fail or shrink: on average, a bit fewer than 50% of new businesses survive five years. In 2016, for example, 49% of firms created in 2011 remained in existence.

Most of the new businesses created each year are service-based businesses in high-turnover sectors. Think about the nearest corner intersection or strip mall in your area: while anchored by national brands or chains, several businesses enter (and exit) in any given year. It might be a restaurant, a retail store, or another similar consumer-facing business.

Not all new businesses make equal contributions to jobs and productivity—of those that survive, only a small subset become the fast-growing companies that have a disproportionate economic impact. These high-growth firms are more likely to be young so, every year, new businesses plus high-growth (young) firms “account for 70% of firm-level gross job creation on average.”⁶³ Yet the median firm in the economy—no matter the size, age, sector, or geography—does not grow in terms of adding new jobs. Net employment growth comes from a fraction of high-growth firms, resulting in what researchers call

⁶³Ryan Decker, John Haltiwanger, Ron S. Jarmin, and Javier Miranda, “Where Has All the Skewness Gone? The Decline in High-Growth (Young) Firms in the U.S.,” *European Economic Review* 86 (2016): 4–23.

“positive skewness” in job growth across American firms.⁶⁴ The median firm (the 50th percentile) doesn’t grow, but high-growth firms (for example, those at the 90th percentile) add thousands of new jobs. The difference between those two points on the spectrum—the 90–50 differential—is a thus a key indicator of business dynamism that some economists look at.

In addition to job creation, new, young, and growing businesses also contribute to overall productivity in the economy. High-growth young firms are “relatively more innovative and productive, so their rapid growth contribute[s] positively to productivity growth as more resources [are] shifted to these growing firms.”⁶⁵

Faltering Dynamism

What all these findings add up to, then, is the empirically established importance of business dynamism to economic growth and productivity.

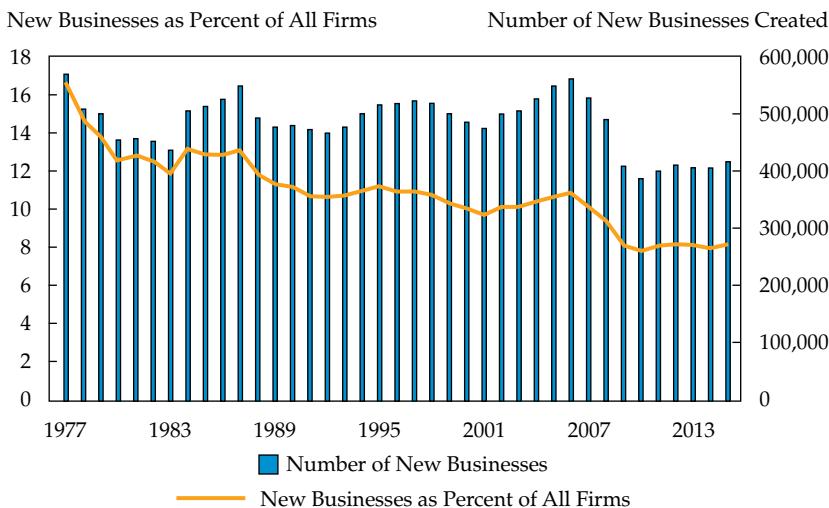
It’s a problem, then, that across nearly every indicator, business dynamism is waning in the United States. New business creation has fallen across the board. High-growth firms are adding fewer jobs and less productivity enhancement. And fewer geographic regions are benefiting from the economic contributions of new and fast-growing businesses. The latest research calls into question America’s continuing self-identify as a startup nation and the future role of entrepreneurship in economic growth.

As **Figure 1** and **Table 1** show, overall business creation in the United States has declined in two respects, volume and pace. The annual number of new businesses, after holding roughly steady for 30 years, fell markedly after the 2008–09 recession and, as of 2016, had yet to fully recover. On average, from 1977 (the earliest year in the dataset) to 1999, 492,000 new businesses were created each year. From 2000 to 2007, this rose to 512,000 new businesses created on average each year, with 2006 recording the highest number of new businesses (557,000) since 1977.

Even though the volume of business creation rose, however, the pace of business creation fell. Because the overall population of firms was growing steadily, with a roughly similar number of new businesses formed each year, the share of businesses that were new slowly declined. In 1979, for example, the 497,000 new businesses were 14% of the total population of businesses; by

⁶⁴Ryan Decker, John Haltiwanger, Ron S. Jarmin, and Javier Miranda, “Where Has All the Skewness Gone? The Decline in High-Growth (Young) Firms in the U.S.,” *European Economic Review* 86 (2016): 4–23.

⁶⁵Ryan Decker, John Haltiwanger, Ron S. Jarmin, and Javier Miranda, “Where Has All the Skewness Gone? The Decline in High-Growth (Young) Firms in the U.S.,” *European Economic Review* 86 (2016): 4–23.

Figure 1. New Business Trends in the United States, 1977–2016

Note: Calculated from Business Dynamics Statistics, US Census Bureau.

2006, the 557,000 new businesses were 11% of the total population. Annual new business creation, from a rate perspective, did not keep pace with the overall growth of businesses. Even though more businesses were created per year between 2000 and 2007 than between 1977 and 1999, the new business share fell because there were nearly 1 million more businesses overall (on average) during the latter period.

In 2008, both the volume and the pace of new business creation fell sharply. From the 2006 peak, new business creation fell for four consecutive years, and was 31% lower in 2010. From 1977 to 2007, new businesses accounted for, on average, 11% of all firms. From 2008 to 2016, they were on average only 8% of all firms. New business creation has fallen before, at least during the time period covered (see **Table 2**).

Table 1. New Business Creation in United States

Time Period	Average # New Businesses Created
1977–2016	479,407
1977–2007	497,883
2000–2007	512,409
2008–2016	415,769

Source: Business Dynamics Statistics, US Census Bureau.

Table 2. Recessionary Declines in New Business Creation

Peak-Trough	% Decline in New Business Creation
1977–1983	–23
1987–1992	–15
1997–2001	–9
2006–2010	–31

Source: Business Dynamics Statistics, US Census Bureau.

Yet the 2006 to 2010 cycle of peak-to-trough new business creation represented the steepest fall in the last 40 years. Most strikingly, new business creation had still failed to rebound by 2016—it appeared to have hit a “new normal” of just over 400,000 new businesses created per year, 17% lower than the annual average between 1977 and 2007. New business creation hit a bottom and . . . just stayed there.

Researchers have labeled this “persistent and widespread collapse in startup rates and the subsequent aging of US businesses as the startup deficit.”⁶⁶ There has been an “unmistakable shift in U.S. firm dynamics since the late 1970s.”⁶⁷ While few people would likely think of the late 1970s as the American entrepreneurial peak, this is evidently what the data tell us. Since then, the entry and exit of firms has fallen, and the pace of “job churn” has also subsided. Economic dynamism, it appears, has gone into quiescence.

Is this such a bad thing? In the 1990s, the retail sector saw a steep decline in dynamism as national chains, led by Walmart, expanded rapidly and drove out thousands of independent retail businesses. This helped lead a historic productivity surge across the entire economy. Perhaps something similar has been happening over the last several years? An intriguing line of research has tried to disentangle this type of quantity-quality question: Even if the quantity of new business creation and dynamism, in the aggregate, has been receding, perhaps quality has not?

There has, after all, been apparent growth in startup activity in areas like agtech (agricultural technology), fintech (financial technology), and even construction tech. Some of the largest, lowest-productivity sectors of the US economy (like construction) have come under assault from tech startups.

⁶⁶Titan Alon, David Berger, Robert Dent, and Benjamin Pugsley, “Older and Slower: The Startup Deficit’s Lasting Effects on Aggregate Productivity Growth” (NBER Working Paper 23875, September 2017).

⁶⁷Fatih Karahan, Benjamin Pugsley, and Aysegul Sahin, “Demographic Origins of the Startup Deficit,” Federal Reserve Bank of New York, May 2018, https://www.newyorkfed.org/medialibrary/media/research/economists/sahin/startpop_latest.pdf?la=en.

Venture capital investment has boomed. Guzman and Stern do indeed find in their analysis that “entrepreneurial quality” has not diminished in the United States even as overall quantity has fallen.⁶⁸ This lends empirical support to the theoretical model of Acemoglu et al. (2017), that young firms are more likely to be innovators. Even reduced aggregate entry (perhaps in retail and services) could coexist with greater innovation (quality) given the higher probability that young firms will be “high type” innovators (in the terminology of Acemoglu et al.).⁶⁹

It may be the case, then, that while aggregate business creation has declined and the country faces an overall “startup deficit,” there could be a substantial increase in entrepreneurial quality underneath the aggregate trends, with more and more tech-driven startups. Overall, this would look like fewer new businesses, but more high-potential dynamism at a micro level in specific sectors and, potentially, geographic regions. If true, then maybe we don’t need to worry about reduced business dynamism overall: Maybe the United States can be a startup nation and have reduced dynamism.

Unfortunately, this does not appear to be the case. Recall the 90–50 differential in employment growth among firms: The median firm barely grows in any year, but the fast-growing firms at the 90th percentile grow extremely rapidly. This “skewness” drives macroeconomic gains in job creation and productivity. But since 2000, the traditional skewness of the American economy has dropped precipitously.⁷⁰ There have been fewer high-growth firms overall and a slower growth rate among those at the 90th percentile. The “disappearance” and “thinning out” of high-growth firms “has substantial macroeconomic effects, lowering aggregate output by 4.5%.”⁷¹

Worse, these declines have been particularly pronounced in the high-tech sector, historically the most dynamic part of the U.S. economy. In economic terms, the high-tech sector “had amongst the highest levels of skewness in

⁶⁸Jorge Guzman and Scott Stern, “The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 15 US States, 1988–2014” (NBER Working Paper 22095, March 2016), <http://www.nber.org/papers/w22095>.

⁶⁹Daron Acemoglu, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William R. Kerr, “Innovation, Reallocation, and Growth” (NBER Working Paper 18993, November 2017), <http://www.nber.org/papers/w18993>.

⁷⁰Ryan Decker, John Haltiwanger, Ron S. Jarmin, and Javier Miranda, “Where Has All the Skewness Gone? The Decline in High-Growth (Young) Firms in the U.S.,” *European Economic Review* 86 (2016): 4–23.

⁷¹Benjamin Pugsley, Petr Sedlacek, and Vincent Sterk, “Disappearing Gazelles: New Evidence from Administrative Data,” VoxEU.org, May 11, 2018, <https://voxeu.org/article/disappearing-gazelles-new-evidence-administrative-data>.

the 1980s and 1990s.”⁷² That is, extremely fast-growing (young) high-tech firms were like rocket boosters to the U.S. economy in the 1980s and 1990s. Yet since 2000, this skewness “has largely been eliminated … Something has happened to the incentives or the ability to be a high-growth firm in the high-tech sector.”⁷³

Impact of Dynamism on Productivity— and Geographic Variation

These trends have major implications for overall productivity and growth. In general, dynamism among firms has historically been a major contributor to greater productivity. New firms enter and young firms challenge existing companies; growing firms put resources to greater economic use than non-growing firms; and, unproductive firms shrink and go out of business. As has already been implied, a reduction in these micro elements of economic dynamism should therefore be expected to drag down macroeconomic progress.

That is precisely what recent research has found: The “startup deficit,” the “reallocation” of resources toward older firms, and the decline in firm dynamism have “reduced aggregate productivity by a little more than 4%, roughly 0.12 percentage points per year.”⁷⁴ In the context of productivity growth, where growth of 2% per year is considered strong, that is a significant drag. Overall output per worker has also been lowered.⁷⁵

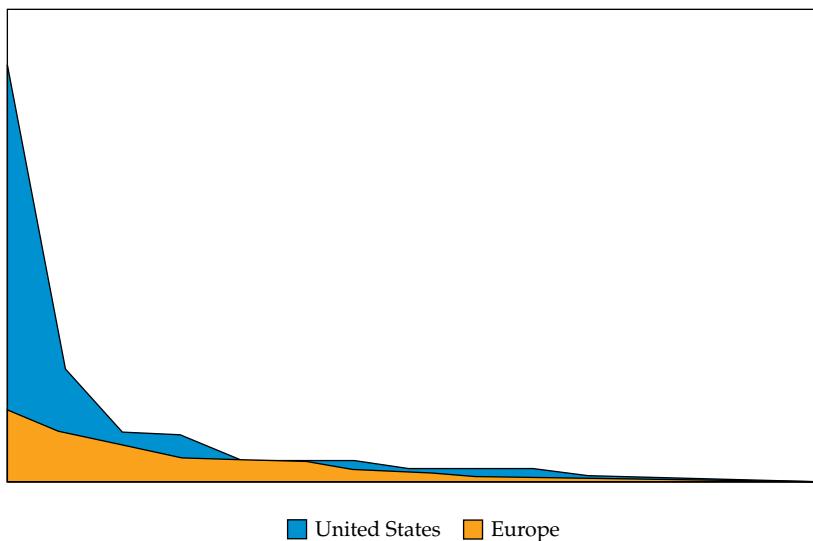
Macro reductions in output and productivity have been aggravated by increasing geographic concentration of the rewards, or economic value, of business dynamism. Reduced dynamism has led to a shrinking economic pie, shared among fewer and fewer places. Analysis by the Economic Innovation Group (EIG) has shown that fewer and fewer metro areas have experienced growth in firm entry. Between 1983 and 1987, 29 metro areas accounted for 50% of the national net increase in new businesses. By the 2010–14 period,

⁷²Ryan Decker, John Haltiwanger, Ron S. Jarmin, and Javier Miranda, “Where Has All the Skewness Gone? The Decline in High-Growth (Young) Firms in the U.S.,” *European Economic Review* 86 (2016): 4–23.

⁷³Ryan Decker, John Haltiwanger, Ron S. Jarmin, and Javier Miranda, “Where Has All the Skewness Gone? The Decline in High-Growth (Young) Firms in the U.S.,” *European Economic Review* 86 (2016): 4–23.

⁷⁴Titan Alon, David Berger, Robert Dent, and Benjamin Pugsley, “Older and Slower: The Startup Deficit’s Lasting Effects on Aggregate Productivity Growth” (NBER Working Paper 23875, September 2017).

⁷⁵Niklas Engbom, “Firm and Worker Dynamics in an Aging Labor Market” (working paper, November 2017), <https://www.niklasengbom.com/wp-content/uploads/JMP.pdf>.

Figure 2. Economic Value Created by Startups Is Highly Concentrated

Note: Ecosystem value (valuations + exits) by ecosystem (geographic region).

Source: Startup Genome.

only five metro areas accounted for that 50%.⁷⁶ Among the “high-type” or “high-quality” startups—those that are technology-based and seeking to grow—the United States has experienced a fairly dramatic skew in how the rewards are distributed (see **Figure 2**).

This trend presents another challenge for those concerned with American entrepreneurship: Business dynamism is not only declining but also narrowing to a smaller geographic distribution.

Startup Fever

How can all this be true? Irrespective of national trends in business creation—and the disputes over quantity versus quality—what is not in doubt is that there has been an explosion of interest and attention in entrepreneurship. The number of support organizations and activities has skyrocketed: accelerators, incubators, pitch competitions, coworking spaces, and so on. The United States—most cities and states, and pretty much every other country as well—has caught what I call “Startup Fever.” Everyone wants more startups, wants to better help those startups, and is creating support systems for startups. You can barely walk down the street in Cedar Rapids or Chattanooga or Reno

⁷⁶Economic Innovation Group, “Dynamism in Retreat,” <https://eig.org/dynamism>.

without encountering “Startup [Fill-in-the-Blank].” This is a global story: Other countries have also experienced declining dynamism in the midst of Startup Fever.⁷⁷

In one sense, it may not be difficult to reconcile these phenomena. Perhaps a person in Kansas City who may once have started a new retail store has today started an e-commerce startup that is more technologically driven and can achieve greater scale through the internet. Maybe, as a consequence, there are three fewer people who would have started retail stores and instead work for that startup or, more likely, at the warehouse distribution center which handles shipping for e-commerce startups.⁷⁸ The e-commerce startup likely participated in a local entrepreneurship program such as an accelerator or pitch competition or university program.

Yet even as Startup Fever means we see more attention paid to entrepreneurship and startups across the country, there are strong headwinds. The biggest is demographic: Recent research has found that the slow growth of the US labor force, driven in part by an aging population, explains most of the startup deficit.⁷⁹ This confirms an explanation posited a decade ago.⁸⁰ An exciting new startup support organization or entrepreneurship program in Oklahoma City or Cleveland may not have much of an effect if local demographic trends are heading in the other direction. One consequence of Startup Fever thus could be delusional economic thinking.

We need to come to terms with the reality of demographic effects, which means coming to terms with data limitations (we have detailed firm dynamic data only since the late 1970s). This, in turn, means coming to terms with what we don’t know. We don’t really know what American entrepreneurial firm dynamics were like prior to the late 1970s. We have a few indicators, based on Census enterprise data from the 1950s and 1960s.⁸¹ But we really

⁷⁷Gert Bijnens and Jozek Konings, “Declining Business Dynamism and Information Technology,” VoxEU.org, July 19, 2018, <https://voxeu.org/article/declining-business-dynamism-and-information-technology>.

⁷⁸Michael Mandel, “How eCommerce Creates Jobs and Reduces Income Inequality,” Progressive Policy Institute, September 2017, http://www.progressivepolicy.org/wp-content/uploads/2017/09/PPI_ECommerceInequality-final.pdf.

⁷⁹Fatih Karahan, Benjamin Pugsley, and Aysegul Sahin, “Demographic Origins of the Startup Deficit,” Federal Reserve Bank of New York, May 2018, https://www.newyorkfed.org/medialibrary/media/research/economists/sahin/startpop_latest.pdf?la=en.

⁸⁰Dane Stangler and Paul Kedrosky, “Exploring Firm Formation: Why Is the Number of New Firms Constant?” Kauffman Foundation Research Series on Firm Formation and Economic Growth, January 2010, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1585380.

⁸¹Dane Stangler and Paul Kedrosky, “Exploring Firm Formation: Why Is the Number of New Firms Constant?” Kauffman Foundation Research Series on Firm Formation and Economic Growth, January 2010, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1585380.

don't have a good indication one way or the other of how unique or representative the last 40 years are in terms of firm dynamics. In 2010, we wrote that it's possible "that the period from the late 1970s to the twenty-first century could be unique in both the level and pace of firm formation, perhaps reflecting technology or other broad changes."⁸² We need to get over the idea that what shaped the last 40 years of firm dynamics—especially new business creation—must be what determines the next 40 years.

There is hope in demographic trends. In fact, if demographic change has been a principal driver of firm dynamics over the last few decades, then we might be on the cusp of an entrepreneurial boom. The next two decades will see more people in the peak age range for starting a business than at any time in US history. Even as the number of Americans over age 65 increases rapidly, the largest age group will remain those between the ages of 25 and 44. This is the prime age for business creation.

While we may have overinvested today in creating support mechanisms for entrepreneurs and startups—resulting in Startup Fever—it's possible that this emerging infrastructure is what enables us to overcome demographic headwinds and declining dynamism and confirm America's claim to be a startup nation.

⁸²Dane Stangler and Paul Kedrosky, "Exploring Firm Formation: Why Is the Number of New Firms Constant?" Kauffman Foundation Research Series on Firm Formation and Economic Growth, January 2010, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1585380.

Understanding the Decline of US Manufacturing Employment

Susan N. Houseman

Vice-President and Director of Research, Upjohn Institute for Employment Research

The manufacturing sector experienced a precipitous and historically unprecedented decline in employment in the 2000s, which coincided with a surge in imports, weak growth in exports, and a yawning trade deficit. The plight of US manufacturing featured prominently in the 2016 presidential election, with candidates Donald Trump and Bernie Sanders arguing that globalization had severely damaged US factories. This argument resonated in many American communities and may have played a role in the election of President Trump. Making good on campaign promises, the president pulled out of the Trans-Pacific Partnership Agreement, proposed renegotiating the North American Free Trade Agreement, and has begun levying tariffs on imports, raising concerns about a trade war.

Countering this view, many economists, policymakers, and pundits cite manufacturing output and productivity statistics to assert that American manufacturing has never been stronger. They point out that although manufacturing employment had been relatively stable before 2000, its share of US employment had been in decline for decades. Often making analogies to the agricultural sector, they contend that automation, not globalization, largely explains manufacturing's relative employment declines and steep job losses in recent years.⁸³

This perspective often is presented as the consensus view among economists and taken as fact in media reports. Typical is a *New York Times* article published in late 2016 in which reporter Binyamin Appelbaum asserts, "From an economic perspective . . . there can be no revival of American manufacturing, because there has been no collapse. Because of automation, there are

Reprinted from the Upjohn Institute Working Paper 18-287, January 2018, with the permission of the publisher, the Upjohn Institute for Employment Research.

⁸³See, for example, Lawrence and Edwards (2013), Muro (2016), and DeLong (2017). Atkinson et al. (2012, pp. 27–28) includes a long list of notable economists and policymakers who subscribe to this view.

far fewer jobs in factories. But the value of stuff made in America reached a record high in the first quarter of 2016, even after adjusting for inflation.”

Regardless of whether the view represents a consensus, it reflects a misreading of the data and research evidence. The apparently robust growth in manufacturing inflation-adjusted (real) output and productivity are driven by a relatively small sector—computers and electronic products, which account for only about 13% of added value in manufacturing. Without the computer and electronic products industry (hereafter referred to simply as “the computer industry”), real value-added or GDP growth in manufacturing was less than half that of the private-sector average from 1979 to 2000, and only 12% in the 2000s. And without the computer industry, manufacturing labor productivity generally has been no higher or only somewhat higher than that of the private sector.

The computer industry, in turn, is an outlier and statistical anomaly. Its extraordinary output and productivity growth reflect the way statistical agencies account for improvements in selected products produced in this industry, particularly computers and semiconductors. Rapid productivity growth in this industry—and by extension the above-average productivity growth in the manufacturing sector—has little to do with automation of the production process. Nor is extraordinary real output and productivity growth an indicator of the competitiveness of domestic manufacturing in the computer industry; rather, the locus of production of the industry’s core products has shifted to Asia.

Manufacturing’s declining employment share has mirrored its declining share of output (nominal GDP) and to a large degree reflects the fact that in most manufacturing industries, there has been relatively little growth in the amount of goods made in American factories for the past 40 years. The recent precipitous decline in manufacturing employment is a distinct phenomenon, and a growing body of research examines the extent to which international trade can explain it. Although none of the studies comprehensively examine the various mechanisms by which trade and the broader forces of globalization may impact employment, collectively they find that trade has played a significant role in the collapse of US manufacturing employment in the 2000s. In contrast, research to date finds little support for the hypothesis that automation was responsible for the sudden decline.

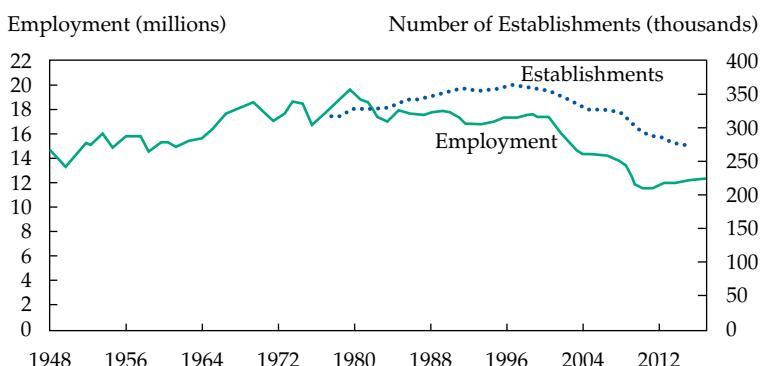
In the remainder of the article, I elaborate on these points. I close with a brief discussion of the consequences of the large job losses in manufacturing for workers and regional economies and consider lessons for policy.

The Collapse of Manufacturing Employment in the 2000s

Figure 1 depicts employment in the manufacturing sector from 1947 to 2016 and the number of manufacturing establishments from 1977 to 2014.⁸⁴ Manufacturing employment trended upward in the years following World War II, peaking at over 19 million in 1979. From 1979 to 1989, the year of the next business cycle peak, manufacturing shed 1.4 million jobs, or 7.4% of its base. The job losses were concentrated in the primary metals and textile and apparel industries. The oil price hikes of the 1970s and early 1980s dampened demand for steel at a time when developing countries were expanding capacity. The resulting excess global capacity led to downsizing in the United States and other advanced economies. The declines in apparel and textiles reflected the shift in production in these labor-intensive industries to developing countries. Employment in manufacturing was relatively stable in the 1990s. Although measured employment declined by about 700,000, or 4%, from 1989 to 2000, the net decline in jobs can be entirely explained by the outsourcing of tasks previously done in-house (Dey, Houseman, and Polivka 2012, 2017).

The precipitous decline in manufacturing employment in the 2000s is historically unprecedented. Between the business cycle peaks of 2000 and 2007, manufacturing employment declined by 3.4 million, or 20%.

Figure 1. Manufacturing Employment and Number of Establishments



Notes: Data on employment are from the Bureau of Economic Analysis. Data on number of establishments are from the Census Bureau's Business Dynamics Statistics.

⁸⁴The Bureau of Economic Analysis recently constructed a consistent time series for industries or sectors from 1947 to the present for data on employment (breakdowns for manufacturing industries since 1977), nominal and real GDP, and GDP price deflators. Most of the analyses in this paper make use of this consistent time series. The number of manufacturing establishments plotted in Figure 1 comes from the Census Bureau's Business Dynamics Statistics.

Although employment in manufacturing, a cyclically sensitive sector, often drops sharply during recessions, the early 2000s marked the first period in which employment in the sector did not entirely or largely recover during the subsequent expansion. Manufacturing employment was hard-hit again during the Great Recession of 2008–2009, rebounding only slightly during the ensuing recovery. From 2007 to 2016, manufacturing employment declined on net by 1.5 million. In total, since 2000, manufacturing employment has fallen by nearly 5 million, or by over 28%. Unlike the declines experienced in the 1980s, the job losses have been broad-based, affecting all industries. Widespread plant closures accompanied the employment declines. As shown in Figure 1, from 2000 to 2014, the number of manufacturing establishments dropped by more than 78,000, a 22% decline.

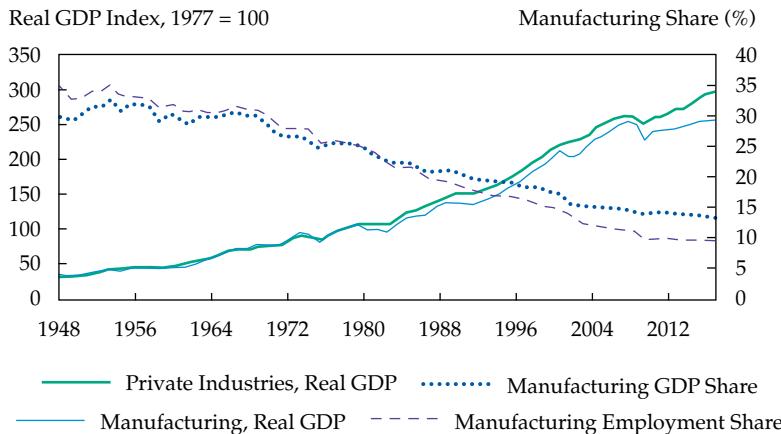
Not only was the sharp decline in manufacturing employment historically unprecedented in the United States, the magnitude of the decline was unique among the world's leading manufacturing economies, according to an analysis by the US International Trade Commission (Benedetto 2018). Over the 1998–2014 period, manufacturing employment significantly expanded in China and in South Korea. Although manufacturing employment shrank by almost 9% in Germany during this period, the drop was far less than in the United States and, Benedetto notes, was accompanied by a 4.8% decrease in the German working-age population. Among the five leading manufacturing economies, only Japan experienced a similar percentage decline in manufacturing employment as the United States, but its working-age population declined by over 9% over the period—in contrast to the United States, where the working-age population grew by more than 16%.

The Puzzle

Reflecting stable or declining employment in the manufacturing sector, the share of US private-sector employment in manufacturing has dropped steadily, and relative declines have been particularly prominent since the 1980s. Manufacturing employment as a share of private-sector employment peaked at 35% in 1953; by 2016, that share had fallen to just under 10%. Manufacturing's share of private-sector GDP has experienced a parallel decline: Manufacturing's contribution to private-sector GDP peaked at 33% in 1953, and by 2016 its share was just 13%. The trends in these shares are depicted in the right scale of **Figure 2**. Together, they suggest that performance in the manufacturing sector has been weak relative to the rest of the economy.

Figure 2 also depicts indices of real GDP for the private sector and for manufacturing (left scale). Although manufacturing output is more cyclically sensitive than the average for the private sector, real GDP growth in

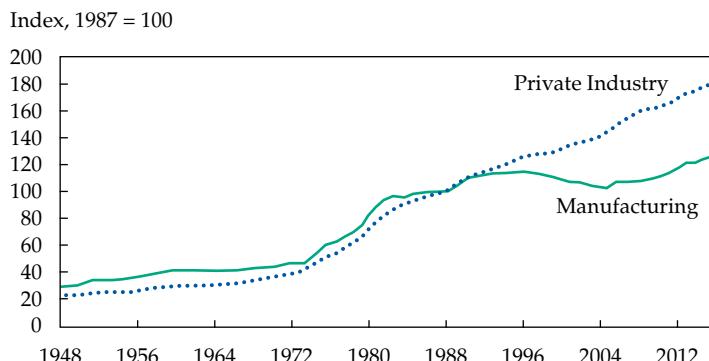
Figure 2. Manufacturing and Private Industry Real GDP; Manufacturing GDP and Employment Shares



manufacturing has largely kept pace with that of the private sector overall. This fact is somewhat paradoxical in view of manufacturing's declining employment and GDP shares. Only since the Great Recession has real GDP growth been considerably slower in manufacturing than in the aggregate economy.

If real GDP growth for manufacturing has kept pace with real GDP growth in the aggregate economy yet manufacturing's share of private-sector GDP is falling, then it must be the case that the average price growth of manufactured goods has been slower than the average price growth for the goods and services produced in the economy. **Figure 3**, which displays an

Figure 3. GDP Price Deflators, Private Industry, and Manufacturing



Note: Data are from the Bureau of Economic Analysis.

index of GDP price deflators for manufacturing and for the private sector, confirms this pattern. The slower growth in price deflators for manufacturing is evident since about 1980. In addition, if real GDP growth has kept pace with real GDP growth in the aggregate private sector yet manufacturing's share of private-sector employment is falling, it follows that labor productivity growth is higher in manufacturing than the average for the private sector.

Reconciling Manufacturing's Declining Shares with Robust Output Growth: The Prevailing Narrative

Manufacturing's declining share of private-sector employment results because manufacturing employment is growing more slowly than the aggregate private-sector employment. Using the fact that labor productivity is defined as output per unit labor, these differential growth rates can be expressed by the following identity:

$$\Delta \ln(L_T) - \Delta \ln(L_M) = [\Delta \ln(GDP_T) - \Delta \ln(GDP_M)] - [\Delta \ln(Prd_T) - \Delta \ln(Prd_M)].$$

In an accounting sense, the difference in the growth rates of labor employed in the aggregate private sector and in manufacturing (L_T and L_M) is equal to the difference in the growth rates of real GDP less the difference in the growth rates of labor productivity. If manufacturing's real GDP growth rate is approximately the same as the average for the private sector, as indicated in Figure 2, then all, or virtually all, of manufacturing's declining employment share is accounted for by higher labor productivity growth.

Although research economists widely recognize that such accounting identities and other descriptive evidence cannot be used to infer causality, many have taken it as strong *prima facie* evidence that higher productivity growth in manufacturing—implicitly or explicitly assumed to reflect automation—has largely caused the relative and absolute declines of manufacturing employment. Even when some role for trade is recognized, it is deemed small, and the decline is taken as inevitable (e.g., DeLong 2017).

Yet productivity growth, which is necessary for improvements in living standards, does not by itself cause employment declines. Productivity growth should lead to higher inflation-adjusted wages, and higher productivity growth in manufacturing should lead to declining prices for manufactured goods relative to other goods and services. This, in turn, stimulates demand for manufactured products. To meet higher demand for their products, manufacturers produce more—potentially fully (or more than fully)

offsetting the adverse effects of higher labor productivity on employment. To reconcile higher manufacturing productivity growth with declining relative and absolute employment, therefore, it must also be the case that consumer demand for manufactured goods is limited and so not very responsive to the declining prices. Analogies are often made to agriculture, where people's food consumption is limited and where mechanization has displaced most farm workers.

An Alternate Reconciliation: Measurement Issues

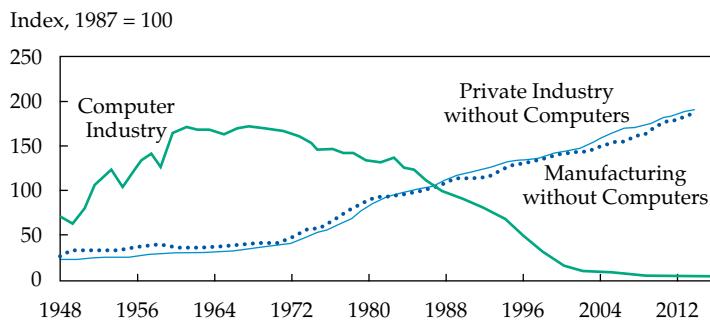
The arguably anomalous patterns depicted in Figure 2—sharply declining manufacturing share of GDP coupled with robust growth in real GDP in manufacturing—imply that price inflation is much lower in manufacturing than in the aggregate economy. One might suppose that there is something unusual about price deflators in manufacturing. Indeed, the strong growth in real manufacturing output is driven by the computer and electronic products industry and reflects the fact that price deflators for certain key products in the industry, namely computers and semiconductors, are adjusted to reflect rapidly improving product quality. For much of the recent past, these price deflators have been sharply declining. Although the computer industry has accounted for less than 15% of added value in manufacturing throughout the period, it has an outsized effect on measured real output and productivity growth in the sector, skewing these statistics and giving a misleading impression of the health of American manufacturing.⁸⁵

Figure 3 depicts price indices used to deflate private industry and manufacturing GDP. **Figure 4** shows price indices for private industry and manufacturing, omitting the computer industry, and for the computer industry by itself.⁸⁶ The price index for the computer industry rises until 1968 and falls thereafter, with particularly steep declines in the 1990s.⁸⁷ Without the

⁸⁵The discussion on the computer industry here and below follows analysis in Houseman, Bartik, and Sturgeon (2015).

⁸⁶In the late 1990s, the BEA began using chained indexes for its real output and price indices in lieu of fixed-weight, constant dollar indices. The chained indices effectively allow the composition of the basket of goods and services to change smoothly over time. Although they avoid biases associated with the old fixed-weight indices, they are computationally more difficult to work with. To back out the computer industry from aggregate price indices and real GDP measures using published data, I employ a Törnqvist index, as described in Houseman, Bartik, and Sturgeon (2015, p. 157).

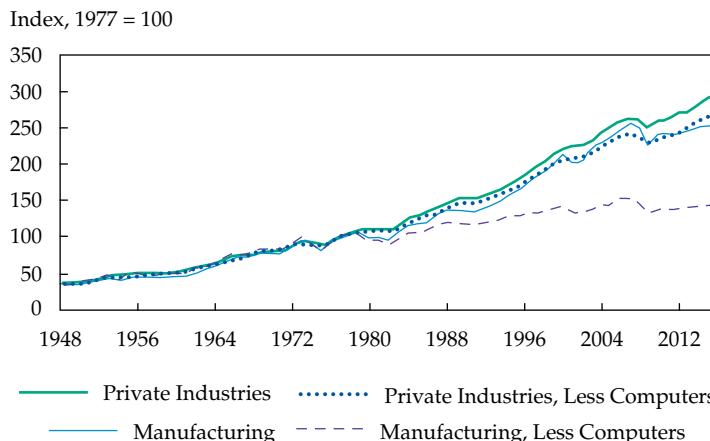
⁸⁷Whelan (2000) notes that in the mid-1980s the Bureau of Economic Analysis began to apply hedonic methods developed by economist Zvi Griliches to construct price deflators for computers. The application of this method, which endeavors to account for the value buyers attach to product quality changes, led to more rapid declines in their price deflators.

Figure 4. GDP Price Deflators in Private Industry and Manufacturing omitting Computer Industry, and in Computer Industry

Note: Author calculations using data from the Bureau of Economic Analysis.

computer industry, the price indices for the private sector and manufacturing display similar trends. Although price inflation for manufacturing without computers has been somewhat lower than the average for the private sector in some years—most notably in the early 1980s and early 2000s—overall the differences are small.

Figure 5 displays indices of real GDP in the private sector and manufacturing, as published and omitting the computer industry. Unsurprisingly, omitting the computer industry has little effect on measured real GDP

Figure 5. Real GDP, Private Industry and Manufacturing, with and without Computer Industry

Note: Author calculations using data from the Bureau of Economic Analysis.

growth in manufacturing prior to the 1980s. The computer industry, however, has had large effects on measured real GDP in manufacturing since then, reflecting the rapid development of semiconductor and computer technology. From 1979 to 2000, measured real GDP growth in manufacturing was 97% of the average for the private sector; when the computer industry is dropped from both series, manufacturing's real GDP growth rate is just 45% that of the private-sector average.

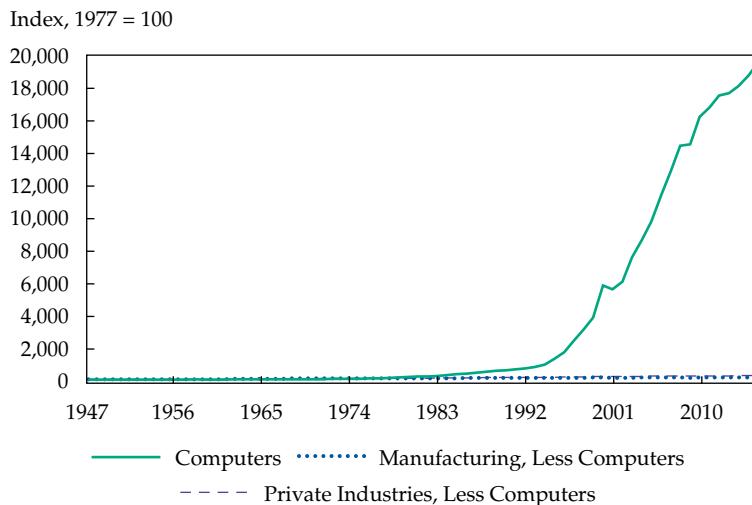
Output growth substantially slowed in both manufacturing and the private sector in the 2000s. In the published series displayed in Figure 2, real output growth in manufacturing was somewhat higher in manufacturing than in the private sector between the business cycle peaks of 2000 and 2007; netting out the computer industry from both series, real output growth in manufacturing was about 60% that in the private sector. Interestingly, without the computer industry, the average rate of real GDP growth in manufacturing was approximately the same over the 2000–07 period, 1.4% per year, as it had been over the 1979–2000 period. While most manufacturing industries experienced lower and in some cases negative real GDP growth in the early 2000s, this was counterbalanced by especially large increases in real GDP growth in the transportation and, to a lesser degree, chemicals industries. I discuss the special case of the motor vehicles industry during this period further below.

Since the Great Recession, real output growth in manufacturing has been noticeably lower than average private-sector real output growth. Just as in prior years rapidly declining computer industry price deflators were responsible for the fact that manufacturing's output growth largely kept pace with that in the aggregate economy, a dramatic slowing of the decline in these price deflators and, correspondingly, of real output growth in the computer industry significantly contributed to the differential growth rates between manufacturing and the aggregate private sector since the last recession.⁸⁸ In published statistics, whereas private-sector output was about 11% higher in 2016 compared to 2007, manufacturing output was approximately the same. Netting out the computer industry, manufacturing output was more than 6% lower in 2016 than in 2007.

Over the entire 2000–16 period, real GDP growth in manufacturing was 63% of the average private-sector growth. Omitting the computer industry from each series, manufacturing's measured real output growth is near zero

⁸⁸Byrne, Oliner, and Sichel (2015) detail the slowdown in the decline of the semiconductor industry's price deflators, and Schmalensee (2018) shows the contribution the computer industry made to manufacturing's lower labor productivity growth during the period.

Figure 6. Real GDP: Private Industry and Manufacturing less Computers



(about 0.2% per year) and just 12% of the average for the private sector in the 2000s.

Figure 6 repeats the series displayed in Figure 5 that omit the computer industry and adds real output growth for the computer industry. The figure illustrates why this industry has such an outsized effect on measured real output growth in manufacturing. Real GDP growth in the computer industry is a different order of magnitude than that for either the private sector or the manufacturing industry series, which omit the computer industries and appear as near horizontal lines along the x -axis because of the different scale needed on the y -axis to accommodate the extraordinary growth in the computer industry. From 1977, the base year in this graph, to 2016 real output in the private sector less computers grew by 169%, real output in manufacturing less computers grew by 45%, while real output in the computer industry increased by 19,257%.

What Explains the Extraordinary Output Growth in the Computer and Semiconductor Industry?

As indicated earlier, the answer to the question of what explains the large and sustained growth in computers and semiconductors lies in the way that the statistical agencies, through the construction of price indices, account for the rapid technological advances in the products produced in this industry. The semiconductors embedded in our electronics are much more powerful

today than they were a decade or even a year ago. Likewise, the computers and related devices that consumers and businesses buy today have much greater functionality than in the past. If, for example, buyers are willing to pay 15% more for a new computer model that boasts greater speed and more memory than last year's model, then 100 of the new computers would be the equivalent of 115 of the previous year's model. The rapid output growth in this industry does not necessarily imply that American factories are producing many more computers, semiconductors, and related products—they may be producing less. Instead, it reflects the fact that the quality of the products produced is better than in the past. The statistical agencies adjust price deflators for other products, such as autos, for changes in quality. However, the effects of quality adjustment in other industries on aggregate statistics, to date, have generally been small compared to those of the computer industry.

It follows that the rapid productivity growth accompanying output growth in the computer industry has little if anything to do with automation: Production of computers and semiconductors has been automated for many years. Rather, rapid productivity growth in the industry—and, by extension, the above-average productivity growth in manufacturing—largely reflects improvements in high-tech products.

Nor is the rapid growth in measured computer and semiconductor output a good indicator of the international competitiveness of domestic manufacturing of these products. As detailed in Houseman, Bartik, and Sturgeon (2015), the locus of production of these products has been shifting to Asia, and the large employment losses in this industry reflect offshoring and foreign competition.

It should be emphasized that the statistical agencies are correct to adjust prices for improvements in product quality. The adjustments, however, can be highly sensitive to methodology and idiosyncratic factors. A change in Intel's pricing strategy for older-generation semiconductors, for example, is partly responsible for the recent slowdown in the rate at which semiconductor price deflators are falling, as explained in Byrne, Oliner, and Sichel (2015). This development has sparked a debate over whether the size of the quality adjustments for the computer and semiconductor industry has been too great or too little.

Such quality adjustment, however, can make the numbers difficult to interpret. Because the computer industry, though small in dollar terms, skews the aggregate manufacturing statistics and has led to much confusion, figures that exclude this industry, as shown in Figure 5, provide a clearer picture of trends in manufacturing output.

Productivity Growth and Interpreting Decompositions That Show Productivity's Contribution to Employment Growth

The computer industry also has a large influence on measured productivity in the manufacturing sector. For various time horizons from 1987 to 2011, Baily and Bosworth (2014) estimate labor and multifactor productivity growth for the private sector, for aggregate manufacturing, and for manufacturing excluding the computer industry. They find that while labor and multifactor productivity growth are considerably higher in manufacturing, when the computer industry is dropped from the calculations, these productivity measures are virtually identical to average productivity growth for the private sector over all time periods examined. As noted from Equation 1, if real GDP growth equals the average growth for the private sector, then productivity growth accounts for all of the relative decline in manufacturing employment. Conversely, if, excluding the computer industry, real GDP growth is lower in manufacturing than in the private sector and labor productivity growth is the same, labor productivity growth can account for none of the relative decline in employment in most of manufacturing.

Since 1977, the Bureau of Economic Analysis has published an industry employment series that is consistent with its industry real and nominal output series. Although employment is a crude measure of labor input because it does not control for differences in hours worked, it allows me to construct the decompositions using Equation 1 for a relatively long-time horizon and show the sensitivity of these decompositions to inclusion of the computer industry. The top panel of **Table 1** decomposes the difference in the average employment growth rate for private industry and manufacturing into the part accounted for by differences in growth rates and the part accounted for by differences in labor productivity. The bottom panel shows this decomposition when the computer industry is omitted from the private sector and manufacturing numbers. From the top panel, over the entire period from 1977 to 2016, average annual employment growth in manufacturing was about 0.025 log points (approximately 2.5%) lower than average employment growth in the private sector. Only 15% of the differential is accounted for by lower output growth in manufacturing, while higher manufacturing labor productivity accounts for 85% of its higher employment growth. When the computer industry is omitted from both series, 61% of the lower manufacturing employment growth is accounted for by manufacturing's lower output growth, and

Table 1. Decomposition of Differences in Private Sector vs. Manufacturing Employment Growth Rates, with and without Computer and Electronic Products Industry, Selected Time Periods

	1977–2016	1979–1989	1989–2000	2000–2007	2007–2016
<i>Including Computer Industry:</i>					
Difference in employment growth rate: private business—manufacturing	0.025	0.029	0.022	0.037	0.019
Share due to GDP growth	0.147	0.195	-0.135	-0.037	0.666
Share due to labor productivity growth	0.853	0.805	1.135	1.037	0.334
<i>Excluding Computer Industry:</i>					
Difference in employment growth rate: private business—manufacturing	0.026	0.032	0.022	0.036	0.019
Share due to GDP growth	0.609	0.478	0.815	0.255	1.020
Share due to labor productivity growth	0.391	0.522	0.185	0.745	-0.020

Notes: The table shows, for various periods, decompositions of the difference in the employment growth rate in the private and manufacturing sectors—with and omitting the computer industry—into the part due to the difference in their real GDP growth and the part due to the difference in their labor productivity growth. Calculations are based on Equation 1 in the text and use data from the Bureau of Economic Analysis.

just 39% by its higher labor productivity growth.⁸⁹ The decompositions are highly sensitive to the inclusion of the computer industry in all subperiods, whose starting and ending years (except for 2016) are business cycle peaks.

The point of this exercise is to show that there is no *prima facie* evidence that productivity growth is primarily responsible for the relative and absolute decline in manufacturing employment. Although such decompositions underlie the narrative that productivity growth, in the form of automation of production, has caused the relative decline in manufacturing employment, they are fraught with measurement problems, and the direction of causality is unclear. If output growth in manufacturing is low relative to the private sector, for instance, it could be because of slower demand growth (domestic or global)

⁸⁹Unlike Baily and Bosworth (2014), I find somewhat higher labor productivity growth in manufacturing compared to the private sector when the computer industry is omitted from both series. The difference partly reflects the fact that Bureau of Economic Analysis made significant revisions to the data, which particularly affected growth in the computer industry, following the publication of the Baily and Bosworth paper.

or the loss of international competitiveness, as evidenced by the growth in the share of imported products or by slow export growth. Some decompositions are embellished to capture changes in output owing to changes in imports and exports. Yet imports and exports must be separately deflated, and existing price indices, particularly import price indices, suffer from well-known biases that understate real import growth (Houseman et al. 2011; Mandel and Carew 2012). In addition, industries are connected by supply chains; imports in one industry will affect demand for inputs in upstream industries, but such effects are not captured in decompositions. Decompositions based on disaggregated industries exacerbate this problem. Job losses owing to trade also may depress domestic demand, but such general equilibrium effects are not captured in these reduced-form accounting identities.

Moreover, labor productivity growth is not synonymous with automation, and measured productivity growth may be simply picking up the effects of international trade and other forces associated with globalization. Given its importance, I elaborate on this last point in the following section.

What Labor Productivity Measures Capture

Labor productivity in an industry or sector is typically defined as added value (the returns to capital and labor) divided by a measure of labor input (hours worked or employment). Labor productivity will increase if processes are automated—i.e., if businesses invest in capital equipment that substitutes for workers in the production process. Measured growth in labor productivity, however, captures many factors besides automation. As already discussed, the strong productivity growth in the manufacturing sector has been driven by productivity growth in the computer industry, which largely stems from product improvements owing to research and development, not from automation of the production process. Although the computer industry has had by far the largest influence on real output and productivity growth in aggregate manufacturing, output and productivity measures in other industries, such as motor vehicles, are significantly affected by quality adjustment of price deflators.

In addition, manufacturers have outsourced many activities previously done in-house, either to domestic or foreign suppliers. If the outsourced activities are primarily done by relatively low-paid, low-value-added workers, or if the outsourced labor is cheaper than the in-house labor, measured labor productivity will mechanically increase. International competition may directly impact measured manufacturing productivity by affecting the composition of products produced and processes used in the United States. The industries and plants within industries most affected by increased competition

from low-wage countries will likely be the most labor-intensive. Similarly, the growth of global supply chains may impact the stages of production done in the United States, affecting labor productivity measures. And exposure to trade can accelerate the adoption of automated processes (Bloom, Draca, and van Reenen 2016; Pierce and Schott 2016). In these cases, there is no simple parsing out of the effects of trade and automation on employment.

A study of plant closures in the early 2000s in the home furniture industry illustrates these forces (Holmes 2011). The making of high-quality wood furniture such as bedroom and dining room furniture, known as casegoods, requires human craftsmanship and does not lend itself to automation. The surge of imports from China and other Asian countries beginning in the late 1990s hit the casegoods industry particularly hard; between 1997 and 2007, a majority of the large casegoods plants shut down, most of the rest downsized, and employment in the industry dropped by half. The upholstery industry was also hard-hit by imports but fared better because of the custom nature of the product and the expense associated with shipping bulky sofas. The US upholstery industry, however, offshored the labor-intensive “cut-and-sew” of upholstery fabric to China in kits, which could be inexpensively shipped. These kits were then stuffed with US-built frames and foam. Holmes investigated two very large plants classified in casegoods that survived the surge of Asian imports. One made ready-to-assemble furniture, thus effectively “outsourcing” the labor-intensive assembly process to the customer; it also had mechanized the stage where finish is applied to the furniture. The other, he discovered, imported all its casegoods from China. The facility engaged in some manufacturing processes for upholstered furniture, but it imported the wood furniture from China and offshored the labor-intensive cut-and-sew work to China. Each of these factors—the shift in the composition of products produced and the stages of production done in the United States as well as a shift toward more mechanized plants—raises measured labor productivity in an industry.

A widely cited Ball State University report illustrates the problem with using accounting identities to draw conclusions about automation’s contribution to manufacturing’s job losses (Hicks and Devaraj 2017). Applying a variant of Equation 1 to manufacturing industries, Hicks and Devaraj conclude that productivity growth accounts for most of the job losses. For example, they claim that from 2000 to 2010, a staggering 3.9 million jobs in the computer and electronics products industry were “not filled due to productivity,” more than five times the number of jobs lost (Table 3). Such a claim is absurd. As noted, the productivity gains in the computer industry largely reflect dramatic improvements in the speed and functionality of computers and related

products, not automation of the production process. While computers sold in 2010 are better than those sold in 2000 (and in a statistical sense a 2010 model counts as more than one 2000 computer model), this does not mean it requires fewer workers to make a 2010 model than a 2000 model.

For the auto industry, Hicks and Devaraj conclude that nearly 600,000 jobs were not filled because of productivity, representing 93% of the industry's job losses over the period. Yet, much of the productivity growth in autos, like computers, reflects product improvements. Since the 1960s, the Bureau of Labor Statistics has adjusted new vehicle price indices for the cost of quality improvements between model years (Williams and Sager 2018). In addition, the development of global supply chains and offshoring of some production during this period, particularly within the NAFTA countries, means that some of the productivity growth likely reflects cost savings and changes in the composition of products produced in the United States. Between the business cycle peaks of 2000 and 2007, the number of vehicles produced in the United States declined at a rate of nearly 5% per year, according to data from the Federal Reserve Board, while real GDP in the motor vehicles industry grew at a rate of about 3.5% per year, according to data from the BEA. The divergent trends in the two quantity measures suggest that adjustment of price deflators for product quality had sizable effects on measured real output growth in the BEA series. The divergent trends are also consistent with offshoring and substantial restructuring of the domestic industry. Automation may well have contributed to job losses in the auto and other industries, but the decompositions in the Hicks and Devaraj report can shed no light on the importance of this factor.

In short, productivity growth does not, *per se*, cause employment declines. Accounting identities and other descriptive evidence cannot be used to draw inferences about the causes of these declines, but once the anomalous effects of the computer industry are excluded, even descriptive statistics provide no *prima facie* evidence that higher rates of automation were primarily responsible for the long-term decline in manufacturing's share of employment. Rather, they suggest that understanding the reasons for the slow growth in manufacturing output—whether from weak growth in domestic demand, strong growth in imports, or weak growth in exports—is critical.⁹⁰

⁹⁰According to BEA data, real growth in domestic consumption of manufactured goods was slower than that for services prior to 2000, consistent with common assertions that faster growth in consumption of services partially contributed to the decline in manufacturing's employment share. Interestingly, real consumption of manufactured goods has outpaced that of services since 2000, which is consistent with consumers' responding to a surge of low-cost imports.

Research on the Causes of Manufacturing's Employment Decline in the 2000s

Accounting identities such as those in Equation 1 are appealing because they appear to provide a simple decomposition of the effects of trade and technology on manufacturing's relative or absolute employment decline. But understanding the causes of the decline in manufacturing employment requires rigorous research. Although such studies are never comprehensive in nature and cannot generate a decomposition of the effects of trade and technology—indeed to some degree developments of the two are interrelated—they can provide insights into the relative importance of the two forces.⁹¹ Recent studies have focused on understanding the causes of the precipitous decline in manufacturing employment in the 2000s.⁹²

That decline coincided with a dramatic widening of the merchandise trade deficit, led by a rise in imports from China. Several studies focus on the effects of Chinese imports on US manufacturing employment. Using regional data at the commuting zone level, Autor, Dorn, and Hanson (2013) estimate that a quarter of the decline in manufacturing employment from 1990 to 2007 is related to the growth of Chinese imports. Pierce and Schott (2016) investigate the effects of granting permanent normal trade relations (PNTR) to China in 2001 and find that manufacturing industries in the United States that were more affected by the change in trade policy, along with their suppliers, experienced larger employment losses. In addition, studies have found sizable adverse effects of Chinese imports on US firm sales, investment, patents, and R&D (Autor et al. 2017; Pierce and Schott 2017), raising larger concerns about the loss of competitiveness of domestic manufacturers and future employment in the sector.

Researchers also have examined the effects of offshoring and exchange rate appreciation on US manufacturing employment. Evidence indicates that offshoring by multinational companies, which have accounted for a disproportionate share of the employment decline, on net has lowered domestic employment (Harrison and McMillan 2011; Boehm, Flaaen, and Pandalai-Nayar 2019). Campbell (2017) estimates that a large, temporary appreciation of the dollar in the early 2000s can explain 1.5 million of the job losses in manufacturing from 1995 to 2008. Moreover, Campbell presents evidence of hysteresis: Job losses from a temporary exchange rate appreciation are not reversed when a currency subsequently depreciates. Economic theory predicts

⁹¹Fort, Pierce, and Schott (2018) also note that research cannot provide such decompositions.

⁹²An earlier version of this paper provides a more detailed discussion of the literature: http://research.upjohn.org/cgi/viewcontent.cgi?article=1305&context=up_workingpapers.

that an appreciation of the dollar could stimulate investments in countries with lower production costs, and firms may be unwilling to write off these sunk-cost investments, even if the dollar depreciates to its original value. Additionally, overseas facilities may become more efficient over time (learning by doing) and thus develop a comparative advantage.

While studies have generally found that factors related to trade have played an important role in the decline of manufacturing employment in the 2000s, studies have failed to uncover a strong relationship between automation of the production process and manufacturing job loss during the period (Acemoglu et al. 2014; Autor, Dorn, and Hanson 2015). Although Acemoglu and Restrepo (2017) estimate that the adoption of robots potentially has large, adverse effects on employment and wages, because the adoption of industrial robots has been limited thus far, it can explain little of the sharp decline in employment that has occurred. Recent studies also have found that the rise of markups since the 1980s and the offshoring of labor-intensive processes (not capital investment) account for the rise of capital share (De Loecker and Eeckhout 2017; Elsby, Hobijn, and Sahin 2013). Such evidence is inconsistent with the hypothesis that a large technology shock caused employment declines and a concomitant rise in the capital share in manufacturing.

The Consequences of Manufacturing Job Losses

With just under 10% of US employment located in the manufacturing sector, some may believe that manufacturing job losses matter little anymore. Through supply chain linkages, however, the manufacturing sector has an outsized effect on the economy. Approximately half of the labor needed in the production of manufactured goods in the United States and other advanced countries is employed outside the manufacturing sector. In addition to job creation effects through these input-output relationships, an increase in employment in the manufacturing sector increases local and national employment by increasing demand: The additional employed manufacturing workers spend more in the economy, creating new jobs. Moretti (2010) estimates that each additional manufacturing job in a city generates 1.6 to 2.5 jobs in local goods and services. Reflecting manufacturing's large spillover effects, research finds that the sudden and large job losses in manufacturing in the 2000s are to a large degree responsible for the weak job growth and poor labor market outcomes among less-educated workers during that decade (Acemoglu et al. 2016; Charles, Hurst, and Notowidigdo 2016).

Moreover, the size of the adverse shock matters for workers' reemployment and earnings and for regional economic outcomes. Workers' long-term earnings losses depend to a large extent on the prevailing local labor market

conditions at the time of the loss; those losing jobs when labor markets are weak suffer larger earnings losses (Jacobson, Lalonde, and Sullivan 1993; von Wachter, Song, and Manchester 2009). Correspondingly, the effects of trade and other adverse economic shocks on regional economies depend critically on the size of the shocks. While local economies can recover from modest setbacks relatively quickly, large adverse shocks can overwhelm a local economy, causing a downward spiral and depressing its economy for decades.⁹³

Conclusion

Two stylized facts underlie the prevailing view that automation largely caused the relative decline and, in the 2000s, the large absolute decline in US manufacturing employment: First, manufacturing real output growth has largely kept pace with that of the aggregate economy for decades, and second, manufacturing labor productivity growth has been considerably higher. These statistics appear to provide a compelling case that domestic manufacturing is strong and that as in agriculture, productivity growth, assumed to reflect automation of production, is largely responsible for the relative and absolute decline in manufacturing employment. Although the size and scope of the decline in employment manufacturing industries in the 2000s was unprecedented, many see it as part of a long-term trend and deem the role of trade small.

That view, I have argued, reflects a misinterpretation of the numbers. First, aggregate manufacturing output and productivity statistics are dominated by the computer industry and mask considerable weakness in most manufacturing industries, where real output growth has been much slower than average private-sector growth since the 1980s and has been anemic or declining since 2000. Second, labor productivity growth is not synonymous with, and is often a poor indicator of, automation. Measures of labor productivity growth may capture many forces besides automation—including improvements in product quality, outsourcing and offshoring, and a changing industry composition owing to international competition. Indeed, the rapid productivity growth in the computer and electronics products industry, and by extension in the manufacturing sector, largely reflects improvements in

⁹³This dynamic is illustrated in Dix-Carneiro and Kovak (2017) in a study of trade liberalization in Brazil in the early 1990s. Regions that initially specialized in industries facing larger tariff cuts experienced prolonged declines in formal sector employment and earnings compared to other regions. Moreover, the impact of tariff changes on the regional economy is persistent and grows over time. The mechanisms, the authors argue, include low labor mobility, slow capital adjustment, and agglomeration economies, which amplify the initial labor demand shock from liberalization.

product quality, not automation of production. In short, the stylized facts, when properly interpreted, do not provide *prima facie* evidence that automation drove the relative and absolute decline in manufacturing employment.

It is difficult to parse out the effects of various factors on manufacturing employment, and research does not provide simple decompositions of the total contribution that trade and the broader forces of globalization make to manufacturing's recent employment decline. Nevertheless, the research evidence points to trade and globalization as the major factor behind the large and swift decline of manufacturing employment in the 2000s. Although manufacturing processes continue to be automated, there is no evidence that the pace of automation in the sector accelerated in the 2000s; if anything, research comes to the opposite conclusion.

Manufacturing still matters, and its decline has serious economic consequences. Reflecting the sector's deep supply chains, manufacturing's plight contributed to the weak employment growth and poor labor market outcomes prevailing during much of the 2000s. Although such large-scale shocks have persistent adverse effects on affected communities and their residents, these costs rarely are fully considered in policy making (Klein, Schuh, and Triest 2003). In addition, because manufacturing accounts for a disproportionate share of R&D, the health of manufacturing industries has important implications for innovation in the economy. The widespread denial of domestic manufacturing's weakness and globalization's role in its employment collapse has inhibited much-needed informed debate over trade policies.

References

Acemoglu, Daron, David Autor, David Dorn, Gordon H. Hanson, and Brendan Price. 2014. "Return of the Solow Paradox? IT, Productivity, and Employment in US Manufacturing." *American Economic Review* 104 (5): 394–99.

———. 2016. "Import Competition and the Great US Employment Sag of the 2000s." *Journal of Labor Economics* 34 (s1, part 2): S141–98.

Acemoglu, Daron, and Pascual Restrepo. 2017. "Robots and Jobs: Evidence from US Labor Markets." NBER Working Paper 23285.

Atkinson, Robert D., Scott M. Andes, Luke A. Stewart, and Stephen J. Ezell. 2012. "Worse Than the Great Depression: What Experts Are Missing About American Manufacturing Decline." Report, Information Technology & Innovation Foundation. <http://www2.itif.org/2012-american-manufacturing-decline.pdf>.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103 (6): 2121–68.

———. 2015. "Untangling Trade and Technology: Evidence from Local Labour Markets." *Economic Journal* 125 (584): 621–46.

Autor, David, David Dorn, Gordon H. Hanson, Gary Pisano, and Pian Shu. 2017. "Foreign Competition and Domestic Innovation: Evidence from U.S. Patents." NBER Working Paper 22879.

Baily, Martin Neil, and Barry P. Bosworth. 2014. "US Manufacturing: Understanding Its Past and Potential Future." *Journal of Economic Perspectives* 28 (1): 3–26.

Benedetto, John. 2018. "Trends in Manufacturing Employment in the Largest Industrialized Economies During 1998–2014." In *U.S. International Trade Commission (USITC) Executive Briefings on Trade* (April). Washington, DC: US International Trade Commission.

Bergsten, C. Fred. 2014. "Addressing Currency Manipulation Through Trade Agreements." Policy Brief (PB) 14-2. Washington, DC: Peterson Institute for International Economics. <https://www.piie.com/publications/policy-briefs/addressing-currency-manipulation-through-trade-agreements>.

Bloom, Nicholas, Mirko Draca, and John Van Reenen. 2016. "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity." *Review of Economic Studies* 83 (1): 87–117.

Boehm, Christoph E., Aaron Flaaen, and Nitya Pandalai-Nayar. 2019. "Multinationals, Offshoring and the Decline of U.S. Manufacturing." NBER Working Paper 25824. <https://www.nber.org/papers/w25824>.

Byrne, David M., Stephen D. Oliner, and Daniel E. Sichel. 2015. "How Fast Are Semiconductor Prices Falling?" NBER Working Paper 21074.

Campbell, Douglas. 2017. "Relative Prices, Hysteresis, and the Decline of American Manufacturing." Unpublished paper. Moscow: New Economic School.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. 2016. "The Masking of the Decline in Manufacturing Employment by the Housing Bubble." *Journal of Economic Perspectives* 30 (2): 179–200.

De Loecker, Jan, and Jan Eeckhout. 2017. "The Rise of Market Power and the Macroeconomic Implications." NBER Working Paper 23687.

DeLong, J. Bradford. 2017. “NAFTA and Other Trade Deals Have Not Gutted American Manufacturing—Period.” Vox Media. <https://www.vox.com/the-big-idea/2017/1/24/14363148/trade-deals-nafta-wto-china-job-loss-trump>.

Dey, Matthew, Susan N. Houseman, and Anne E. Polivka. 2012. “Manufacturers’ Outsourcing to Staffing Services.” *Industrial & Labor Relations Review* 65 (3): 533–59.

———. 2017. “Manufacturers’ Outsourcing to Temporary Help Services: A Research Update.” Economic Working Paper 493. Washington, DC: US Department of Labor Bureau of Labor Statistics, Office of Employment and Unemployment Statistics. <https://www.bls.gov/osmr/research-papers/2017/ec170010.htm>.

Dix-Carneiro, Rafael, and Brian K. Kovak. 2017. “Trade Liberalization and Regional Dynamics.” *American Economic Review* 107 (10): 2908–46.

Elsby, Michael W., Bart Hobijn, and Ayşegül Şahin. 2013. “The Decline of the U.S. Labor Share.” *Brookings Papers on Economic Activity* 2013 (2): 1–63.

Fort, Teresa C., Justin R. Pierce, and Peter K. Schott. 2018. “New Perspectives on the Decline of US Manufacturing Employment.” *Journal of Economic Perspectives* 32 (2): 47–72.

Harrison, Ann, and Margaret McMillan. 2011. “Offshoring Jobs? Multinationals and U.S. Manufacturing Employment.” *Review of Economics and Statistics* 93 (3): 857–75.

Holmes, Thomas J. 2011. “The Case of the Disappearing Large-Employer Manufacturing Plants: Not Much of a Mystery After All.” Economic Policy Paper, Federal Reserve Bank of Minneapolis.

Houseman, Susan N., Timothy J. Bartik, and Timothy Sturgeon. 2015. “Measuring Manufacturing: How the Computer and Semiconductor Industries Affect the Numbers and Perceptions.” In *Measuring Globalization: Better Trade Statistics for Better Policy—Volume I. Biases to Price, Output, and Productivity Statistics from Trade*, edited by Susan N. Houseman and Michael Mandel, 151–194. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Houseman, Susan, Christopher Kurz, Paul Lengermann, and Benjamin Mandel. 2011. “Offshoring Bias in U.S. Manufacturing.” *Journal of Economic Perspectives* 25 (2): 111–32.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan. 1993. "Earnings Losses of Displaced Workers." *American Economic Review* 83 (4): 685–709.

Klein, Michael W., Scott Schuh, and Robert K. Triest. 2003. *Job Creation, Job Destruction, and International Competition*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Lawrence, Robert Z., and Lawrence Edwards. 2013. "U.S. Employment Deindustrialization: Insights from History and the International Experience." Policy Brief (PB) 13-27. Washington, DC: Peterson Institute for International Economics. <https://www.piie.com/sites/default/files/publications/pb/pb13-27.pdf>.

Mandel, Michael, and Diana G. Carew. 2012. "Measuring the Real Impact of Imports on Jobs." Progressive Policy Institute (PPI) Policy Memo (March). Washington, DC: Progressive Policy Institute. https://www.progressivepolicy.org/wp-content/uploads/2012/03/03.2012_Mandel-Carew_Measuring-the-Real-Impact-of-Imports-on-Jobs.pdf.

Moretti, Enrico. 2010. "Local Multipliers." *American Economic Review* 100 (2): 373–77.

Muro, Mark. 2016. "Manufacturing Jobs Aren't Coming Back." *MIT Technology Review* (18 November). <https://www.technologyreview.com/s/602869/manufacturing-jobs-arent-coming-back/>.

Pierce, Justin, and Peter Schott. 2016. "The Surprisingly Swift Decline of US Manufacturing Employment." *American Economic Review* 106 (7): 1632–62.

Pierce, Justin, and Peter Schott. 2017. "Investment Responses to Trade Liberalization: Evidence from U.S. Industries and Plants." NBER Working Paper 24071. <https://www.nber.org/papers/w24071>.

Schmalensee, Richard. 2018. "The Curious Collapse of U.S. Manufacturing Labor Productivity Growth After 2010." Unpublished manuscript. Cambridge, MA: Massachusetts Institute of Technology.

von Wachter, Till, Jae Song, and Joyce Manchester. 2009. "Long-Term Earnings Losses Due to Mass Layoffs During the 1982 Recession: An Analysis Using U.S. Administrative Data from 1974 to 2004." Unpublished manuscript. http://www.econ.ucla.edu/tvwachter/papers/mass_layoffs_1982.pdf.

Whelan, Karl. 2000. “A Guide to the Use of Chain Aggregated NIPA Data.” Washington, DC: Federal Reserve Board.” <https://www.federalreserve.gov/pubs/feds/2000/200035/200035pap.pdf>.

Williams, Brendan, and Erick Sager. 2018. “A New Vehicles Transaction Price Index: Offsetting the Effects of Price Discrimination and Product Cycle Bias with a Year-Over-Year Index.” Paper presented at the Meeting of the Group of Experts on Consumer Price Indexes, Geneva, 7–9 May, by the US Department of Labor Bureau of Labor Statistics. https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/BLS_Poster.pdf.

Two Economies: The Decline of the American Worker and the Rise of Superstars

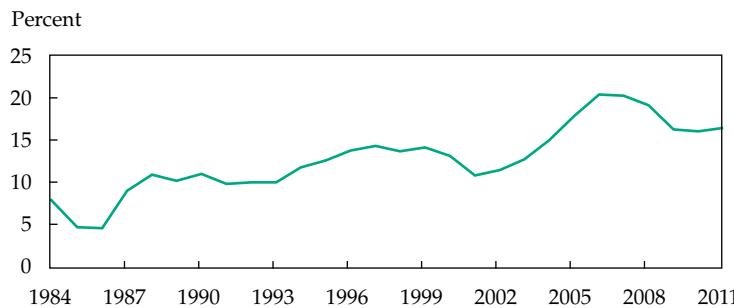
Anton Korinek

Associate Professor of Economics, University of Virginia and Darden School of Business, and Research Associate, NBER

Economic dynamism—growth, change, creative destruction—is imperative for economic prosperity, and it is natural that the question of how to restore economic dynamism ranks high on the agenda of anybody who is concerned with the present malaise of the American economy. However, if we take a closer look, the US economy has actually bifurcated into two economies. On the one hand, there is a highly dynamic growth economy that is inhabited by what I want to call superstars, sometimes dubbed the “1%.” Walking the streets of New York City or San Francisco, one can truly feel the dynamism. On the other hand, a large fraction of the population has been left behind and has not seen any income gains in recent decades. Sometimes this part of the population is dubbed the “99%,” even though the condition probably afflicts more like 75% to 85% of the population; we still have a significant upper-middle class that is doing fine. However, for ease of exposition I will stick to the 1% versus 99% terminology. For the 99%, the economy feels as dismal as the science that studies it. And, given this bifurcation, the problem is not so much restoring American economic dynamism but spreading around economic dynamism so that all Americans will benefit from it.

Let us start with the data. Over the past three decades, we have seen the labor share of income decline from 64% to 58%. We have seen the income of the bottom 90% decline from 67% to 52% of total income while the share of the top 1% has risen from 8% to 19%. The pure profits that firms are earning, which include rents, markups, and so forth, have risen from 2% to 16% according to estimates by Simcha Barkai, an economist at the London Business School. I put together these numbers in what one may call the “superstar share” of the economy, which consists of the labor income of

This article is a background note for a presentation given at the CFA Society New York Conference Restoring American Economic Dynamism on 28 November 2017.

Figure 1. Superstar Share

the top 1% plus the pure rents earned by corporations. The resulting graph is given in **Figure 1**. The superstar share of the economy has risen from approximately 5% to 15% and 20% of our economy.

What are the economic forces behind this superstar phenomenon? In a recent paper titled “Digitization and the Macro-Economics of Superstars,” which I have coauthored with Ding Xuan Ng, we argue that the main force behind the proliferation of superstars is digital innovation, which encompasses advances in the collection, processing, and provision of information.⁹⁴ We view digital innovation as a very broad phenomenon that occurs in almost all sectors of the economy these days because we use computers everywhere. Digital innovation allows firms or entrepreneurs to replace an increasing fraction of the tasks that are required to produce something by digital or information technologies, which differ critically from traditional production technologies. Information technologies have the property that you establish them once and then you can apply them at close-to-zero additional cost to a broad market. A typical example for this is internet companies—Google programs its website once, and then it can be used by millions around the globe. Digital technologies have also become quite widespread on Wall Street—you program a trading algorithm once, and (at least in some cases) it makes little difference if it is used to trade \$100 million or \$100 billion. The superstar phenomenon also takes place in sports, the arts, and music. A similar effect also occurs for franchise owners, which are all about spreading information on best industry practices.

⁹⁴The title of the paper is a play on the title of the celebrated article by Sherwin Rosen, “The Economics of Superstars,” *American Economic Review* 71, no. 5 (1981): 845–58. Rosen was the first to identify the superstar effect. We added the prefix “Macro” to his original title to emphasize that the superstar phenomenon has now become macroeconomically relevant.

The main economic mechanism behind the rise of superstars is that information, and by implication information technology, is (i) *non-rival* but, at the same time, (ii) *excludable*.

(i) What does *non-rival* mean? Most goods are what economists call *rival*, which means that they are used up as they are used. For example, if I use a ton of steel in the production of a car, the steel is used up, my competitor (my rival) can no longer use it. To produce another car, another ton of steel needs to be bought. By contrast, if I use information, it is not used up when being used—I can use the same information again and again, and other people can use it too. What that implies is that a digital innovation can be used to supply a very large market at almost negligible cost. This gives rise to increasing returns, in the sense that every time I produce an additional unit of output, I can spread the fixed cost of producing the digital innovation over a larger number of units produced, lowering the average cost of production.

As a result, digital innovation creates natural monopolies, meaning that it would be natural to produce the information good behind it only once (or, at most, a small number of times) to avoid wasteful duplication. For example, it is natural that we have one dominant search engine, one dominant social media network, one dominant shopping website, etc., because it was very costly to develop their systems, and it would be wasteful and inefficient to create similar systems and companies many times over. Although natural monopolies are particularly important in the information economy, they have existed for a long time in other parts of the economy such as in utilities—it would be wasteful and inefficient for 20 different water companies to build competing water grids to supply a given city; usually we have only a single water company to avoid duplication.

(ii) What does *excludable* mean? If I have invented something, for example, if I have developed a new algorithm or a new computer program, I can exclude others from using it. This can be done using both intellectual property rights or using business secrets. The *excludable* nature of information goods implies that innovators can frequently employ their natural monopoly power to extract large returns. Some of these returns are necessary to finance the development of digital innovations. The information economy requires some rents to keep the system going and finance further innovation. But the amount of rents generated in today's economy is almost certainly in excess of what is necessary to finance further innovation. In other words, there are excess monopoly rents that accrue to the superstars. One could say that digital innovation supercharges the superstar phenomenon.

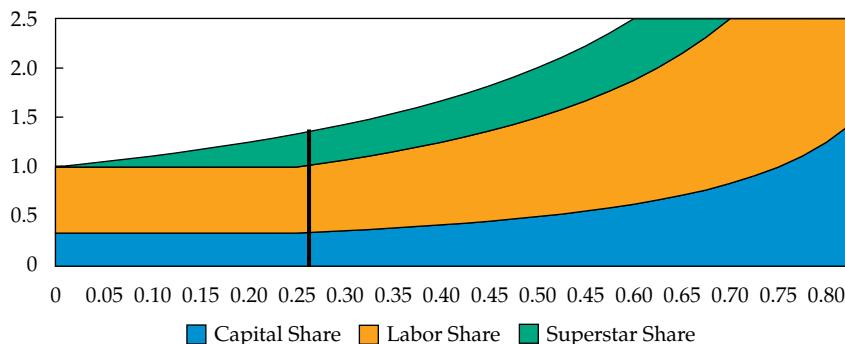
To look at the effects of digital innovation in more depth, let us consider how digital innovation will affect a sector of the economy that originally employed a traditional bricks-and-mortar type of technology, employing labor and capital to produce output. Assume that an entrepreneur introduces a digital innovation in the sector that allows her to substitute for a fraction of the tasks involved in producing output. The first thing to happen is that the entrepreneur will be able outcompete the traditional firms in the sector and take over much of the market, i.e., become a superstar in that sector. The innovator will initially not pass on most of her cost savings but will absorb them as increased markups, generating growing monopoly profits, although part of these are used to cover the costs of the innovation.

As the entrepreneur keeps innovating and automating to further reduce production costs, she will at some point have reached her optimum monopoly markup (i.e., the point where it is profit-maximizing to reduce prices in order to expand the size of the market). She will start to pass on any further cost savings to her customers, who in turn respond by demanding greater and greater quantities of the good. At that point automation is starting to significantly reduce consumer prices and contribute to consumer welfare.

What do these dynamics imply for the demand for labor? At first, the demand for labor declines because the superstar technology allows the entrepreneur to produce more and more with less—this is the traditional *labor-saving* effect of innovation. However, once the entrepreneur reaches the optimal monopoly markup and quantities start to increase, any further gains from progress are shared equally among labor, capital, and superstars. That is the output *scale effect* of innovation.

Figure 2 illustrates these dynamics by plotting the earnings of labor, capital, and superstars as the entrepreneur automates an increasing fraction of

Figure 2. Digital Innovation and Factor Income



the production process. At the very left of the graph, when we start with the status quo where only the traditional technology is used, the returns on capital (blue area at the bottom) and labor (yellow area in the middle) correspond to the traditional capital and labor shares of roughly one-third and two-thirds of the economy. As we move farther to the right in the graph, the fraction of tasks that is automated with digital technologies increases.

At first, the returns to labor and capital are completely stagnant, even though production in the economy rises considerably. Any gains in economic output accrue to the superstars (green area at the top). Once the economy hits the threshold where superstars earn the optimal monopoly markups, they start lowering prices. At that point, the superstar profit share, the labor share, and the capital share remain constant, and the earnings of the three factors grow in tandem. In summary, digital innovation leads to a declining labor share and a rising superstar profit share for some time. But at some point, the superstar profit share is bounded by the optimal monopoly markups in the economy, which limits how much in superstar rents digital innovators can capture. From then on, further digital innovation will be spread broadly around the economy.

Let us now turn to policy implications. An immediate insight of our analysis is that this type of digital innovation introduces monopoly distortions into the economy, which come in two forms. First, given that digital innovators are natural monopolists, there will be too little innovation. Second, there will be too little output produced in those sectors as they charge markups and engage in monopoly pricing. The intuition is that whenever there are markups, the economy is not operating at its efficient level.

I have listed a number of policy remedies in the associated paper, but let me give a brief summary: One natural policy solution is to use, to the extent possible, public investment to finance digital innovation. Digital innovations, like all information goods, are natural public goods, and in an ideal world, they should be financed publicly. In practice, there are some areas where private enterprise is better suited to innovate. However, we should invest as much as possible in basic research, rather than cutting university funding, as is currently being discussed.

A second policy remedy is to investigate whether the factors behind the natural monopolies can be freed up by appropriate policies. For example, if we free up the information that gives big tech firms a monopolistic advantage, such as information about us as consumers or our social networks, we can expect that it will unleash a wave of entrepreneurialism. Imagine you can give any startup that you trust access to your search history, or to your social

network tree, with the click of the button. That would unleash a lot more entrepreneurialism and cut into those monopoly rents.

A third point to emphasize is that in this type of information economy, flows of information matter increasingly for our trade agreements. When people question the fairness of US trade agreements, what they should focus on in our modern times is trade in information goods, not manufactured goods. Information goods are the main comparative advantage of the US and make up an increasing part of our exports, even if our trade statistics don't account for their full value, since many of the gains are recorded in tax havens. Information goods should thus also play a central role in our export strategy.

I would like to thank David Adler and Larry Siegel as well as the participants of the conference for helpful comments on the topic.

Finding Capital in the Stock Market

David Blitzer

Managing Director and Chair of the Index Committee, S&P Dow Jones Indices

Capital, and capital spending, is essential for productivity growth. Traditionally, companies seeking to raise capital came to the stock market. However, times change; companies now raise capital both on and off the market. In addition, there are fewer companies on the stock market than 10 years ago, and the nature of capital itself is changing. These changes began before the financial crisis, and the forces affecting the stock market are echoing throughout the entire economy.

Capital, and especially access to new capital, is an important contributor to productivity. Investing capital for increased productivity may take many forms: physical capital in new plant and equipment, intellectual capital in R&D, or human capital through education.

The Stock Market

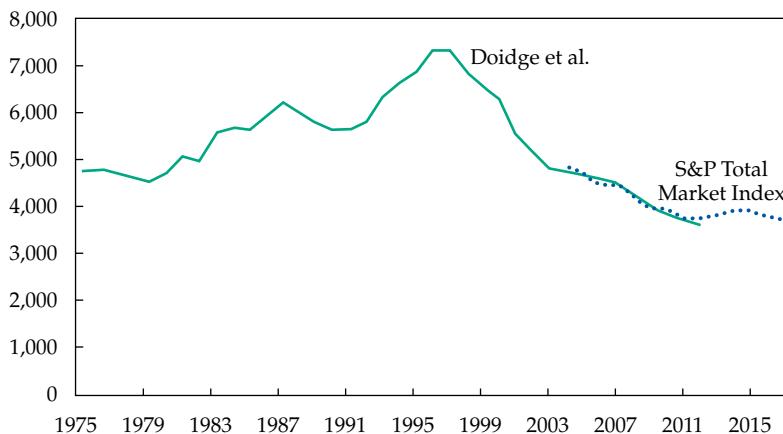
Today the US stock market numbers about 3,800 listed companies, ranging from Apple, valued at around \$900 billion, to the smallest, valued at only a few million dollars.

Ten years ago, there were roughly 4,500 companies in the US market, some 20% more than today. Twenty years ago, the count of companies was even higher, peaking at an all-time high of 7,500 listings.⁹⁵ (**Figure 1**) At 3,800 companies the US remains the largest developed market, followed by Japan and the United Kingdom; China and India, respectively, are the largest emerging markets. Stock markets in other developed countries have also seen their company counts decline, but not by as much as the US (see Figure 1). Emerging markets, at a different stage of market development, are not experiencing the same declines in company listings.

Fewer companies don't mean less wealth. While the number of listed companies is shrinking, the US market's value is growing. In 2017 the S&P 500®

This paper represents my own views, not those of S&P Dow Jones Indices or S&P Global Inc.

⁹⁵Craig Doidge, G. Andrew Karolyi, and Rene M. Stulz, "The U.S. Listing Gap" (Dice Center Working Paper 2015-7, Ohio State University, 2016); S&P Total Market Index, S&P Dow Jones Indices.

Figure 1. Number of Publicly Listed US Corporations

Source: Craig Doidge, G. Andrew Karolyi, and Rene M. Stulz, "The U.S. Listing Gap," June 2016; S&P Total Market Index, S&P Dow Jones Indices.

gained 19.4%. Further, the market at the end of 2017 was worth about 3.8 times its value on March 9, 2009, when the last bear market ended. The increase in value is attributable to rising stock prices, not new capital entering the market. Over the first three quarters of 2017, when the market rose 12.5%, \$1.1 trillion of non-financial corporate equity left the public and private markets.⁹⁶

Today's market is obviously not the market of 10 or 20 years ago. The total number, total value, and median size of companies are all different. Companies today are fewer and larger. The investors' opportunity set—the range of available investments—has shrunk. While the US market is the biggest by count, and 3,800 sounds numerous, in many industries there is actually much less variety in investment opportunities. For example, categories such as large money center banks, integrated oil companies, tobacco producers, retail drug store chains, airlines, gold miners, or beer producers now each have 10 or fewer companies. Moreover, the new industries in the market that didn't exist 10 or 20 years ago aren't replacing those lost in traditional industries: A handful of stocks dominate the internet and new technology industries too.

Capital Flows In and Out

What forces shrunk the market? The market's value changes when investors bid stock prices up or down. In addition, changes occur when capital flows

⁹⁶Financial Accounts of the United States (Z.1), Board of Governors of the Federal Reserve System, Third Quarter 2017, Table F223.

into or out of the market. Capital exits when companies buy back stock or pay dividends. Sources of capital in-flows are initial public offerings (IPOs), secondary offerings or acquisitions paid with cash.

Stock indices measure the value of the market as driven by stock prices. As an example, the S&P Total Market Index has a value of \$28 trillion, which is the sum of the value of all 3,800 companies in the index and the market. The most important factor driving the index's value is the rise and fall of stock prices. However, other factors will affect the market: IPOs add new companies, while share buybacks reduce the number of shares and drain capital out of the market. To correctly calculate value with an index, one must adjust for capital flows from events such as buybacks or IPOs.

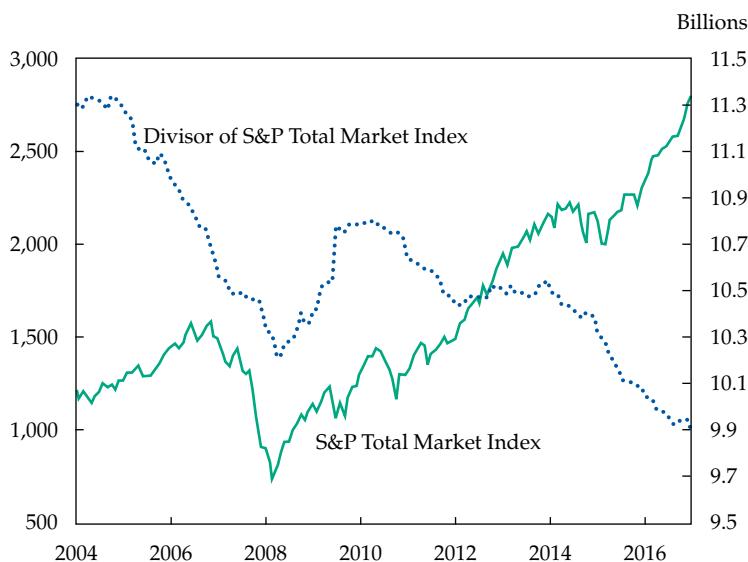
The index is calculated by dividing the total market value of the companies in the index by a scale factor, called the divisor. This divisor has two functions. First, it scales the index to numbers that are easier to handle than \$27.09 billion. Second, if stocks are added or removed, or if there are other capital adjustments, then the divisor is adjusted so that adding a stock doesn't cause the index level to jump. When capital flows into or out of an index, the percentage change of the divisor equals the percentage change of the capital included in the index. If several stocks are added to the index, thus increasing the total index market value by 10%, the divisor is adjusted upward by 10% so that the index level remains the same.

Figure 2 shows the movement of the S&P Total Market Index and its divisor from 2005 to the end of 2017. The index began to fall in October 2007 and bottomed out in March 2009. From there the index rose, with occasional bumps, through the end of 2017. During the initial period after the 2009 bottom the divisor increased as stocks were added to the index and the market. From late 2010 the divisor slid, due to a combination of companies leaving the market, dividend payments and buy backs. The loss in capital from March 2011 to the end of 2017 was 8% of the 2011 level.

A broader measure of capital flows into or out of the stock market is found in the Federal Reserve's Financial Statement of the United States, which details the net issuance of corporate equity. Recently the net issuance has been negative and capital is leaving the market. From the first quarter of 2012 through the third quarter of 2017, \$10.0 trillion of non-financial corporate equity vanished. To provide a sense of scale, non-financial corporate equity totaled \$25.8 trillion at the end of the third quarter of 2017.⁹⁷

The Federal Reserve data show that corporate equity had been declining since the early 1980s then fell sharply in 1987, following the market crash, as

⁹⁷*Financial Accounts of the United States* (Z.1), Board of Governors of the Federal Reserve System, Third Quarter 2017, Tables F223 and L223.

Figure 2. S&P Total Market Index and Divisor

Source: S&P Dow Jones Indices.

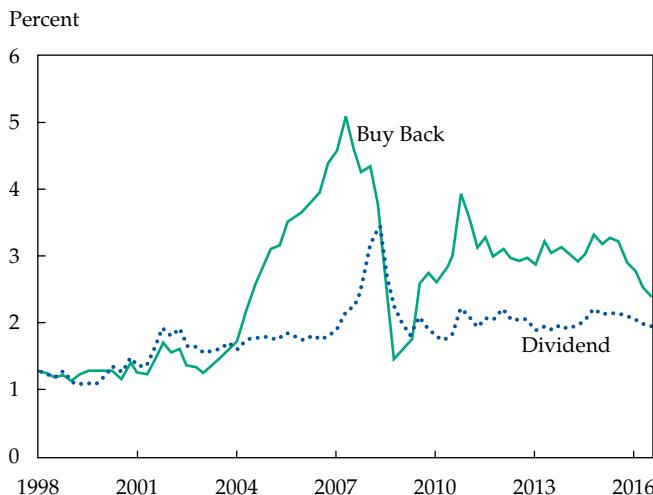
firms sought to support stock prices through buying back stock or increasing dividends.

While buying back stock may have first made the news in 1987, today it is a routine event. Investors like stock buybacks: buybacks return capital to shareholders and raise reported earnings per share. Shareholders desiring to stay invested get to own a slightly larger proportion of the company, and investors seeking income can time the sale of their shares and have the transaction taxed as a capital gain.

Companies usually announce buyback programs in advance, get credit for it in the media, and then make purchases over time. Rarely do shareholders check on the completed size of the buyback. Moreover, a company that reduces the size of a buyback program doesn't get penalized by a reduced stock price the way cutting the dividend would, thus limiting risk of negative consequences to a company choosing to offer a buyback versus a dividend increase. Another factor behind the growth of buybacks is the use of stock grants and options as part of employee compensation. Absent buybacks, companies would face dilution of their equity as employees exercised options or received stock grants.

Data on companies in the S&P 500 show how companies returned capital from the market to investors through buybacks and dividends. Since 2004, buybacks have dominated dividends except for a brief period from the second

Figure 3. Buyback and Dividend Yields for the S&P 500



Source: S&P Dow Jones Indices.

quarter of 2009 to the first quarter of 2010. From 2010 to 2015 the buyback yield (per share value of buybacks divided by the stock price) averaged 3.0%, while the dividend yield over the same period was 2.0%. Since the second quarter of 2016 both the buyback and the dividend yields fell to 2.4% and 1.9%, respectively, in the third quarter of 2017.⁹⁸

If companies were simply retaining and investing a large portion of their earnings, the capital drain from buybacks and dividends wouldn't merit much attention. Instead, as a group, S&P 500 member companies together paid out approximately 97% of their entire earnings as buybacks and dividends from 1999 through 2017.⁹⁹ **Figure 3** shows the buyback and dividend yields for the S&P 500 since 1999.

Companies Moving In and Out

Mergers and acquisitions are one factor in the declining number of public companies. Large or controversial transactions like the recent debate over the AT&T-Time Warner deal make headlines. Behind these well-documented transactions is a fairly steady flow of mergers. Merger activity climbed steadily

⁹⁸“S&P 500 Q3 2017 Buybacks Increase 7.5% to \$129.2 Billion” (press release, December 13, 2017), https://us.spindices.com/documents/index-news-and-announcements/20171213-sp-500-q3-2017-buybacks-press-release.pdf?force_download=true.

⁹⁹Ibid.

from 1985, peaking at the end of the 1990s. When tech stocks collapsed, the number of mergers declined for two years before resuming the upward trend, climbing to a new peak in 2007 when the financial crisis struck. The subsequent drop was short lived, and it bottomed out at a level higher than the 2002 low. Since 2009 merger activity has been rising.

Stepping back from year-to-year variation, over the last 20 years mergers have been a consistent part of the capital markets. The M&A Institute data covering all mergers,¹⁰⁰ not just public companies, shows an average annual value of mergers of \$1.5 trillion over the last 10 years. A narrower source of data is the Hart-Scott-Rodino (HSR) filings with the Federal Trade Commission. The HSR data reveal a pace of 1,000 to 2,000 transactions annually. Moreover, that flow is not interrupted by antitrust actions. Extremely few HSR filings lead to challenges—in 2016 there were only 22 challenges out of 1,800 merger filings.¹⁰¹

Mergers reduce the number of companies listed on the US market. However, mergers do not always drain capital from the markets. In a merger where the acquiring company issues shareholders of the target company new shares for the entire purchase price, there is no loss of capital. If the transaction is a combination of cash and stock, the capital represented by the cash portion will exit the market while the stock portion of the purchase is capital retained in the market.

A merger of two US companies will reduce the number of listings without shrinking the capital in the market. However, cross-border mergers or mergers where the acquirer is a private company do drain capital. Two recent examples are Bayer AG's acquisition of Monsanto and Dell's acquisition of EMC Corp. Bayer is foreign, whereas Dell is private.

One other effect of merger activity, separate from the drop in the total number of US-listed companies, is the increasing concentration across many industries that is resulting in fewer, larger companies. Until recently mergers or increasing concentration provoked little reaction. Now, given the size of certain tech and telecom companies, there is increasing concern about the influence and political power such companies enjoy.

IPOs

The most discussed reason for the shrinking company list is a dearth of IPOs. Consider some of the numbers: From 1980 to 1989 the United States averaged 200 IPOs per year; from 1990 to 1999 that number doubled to just over

¹⁰⁰Institute for Mergers, Acquisitions and Alliances, <https://imaa-institute.org/>.

¹⁰¹Federal Trade Commission, *Hart-Scott-Rodino Annual Reports*, <https://www.ftc.gov/policy-reports/policy-reports/annual-competition-reports>.

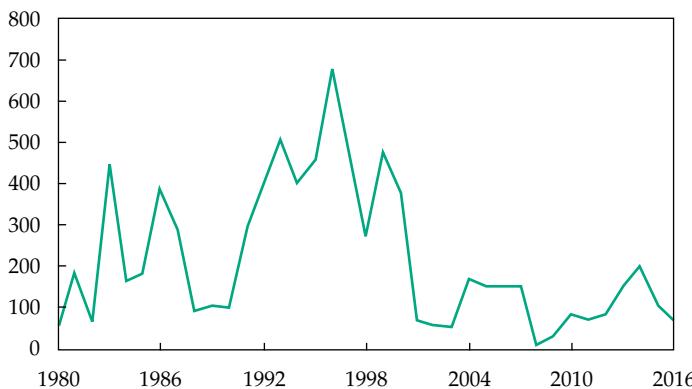
400 per year. In the first decade of the 21st century, IPOs plunged to only 130 per year; and since 2010, they are averaging only 117. Had we maintained the pace of the 1990s for the last 16 years, we would have seen 4,500 more IPOs. Even if only half survived, the market would have been roughly half again as big as it is now (see **Figure 4**).¹⁰²

Two aspects of the IPO tally are worth noting: First, the number of technology IPOs averaged around 30% of the total count from 1980 to about 1995, jumped to over 50% until the tech stock collapse in 2000, and is now back to about 30%. The weak IPO numbers are partly, but not totally, attributable to the 1990s tech boom-bust cycle.

Second, IPOs are changing just as listed companies are changing. The median age of IPOs in the 1990s was 8 years; from 2000 to 2016 it was 11 years. Now startups spend more time as private companies before reaching the stock market. Newly listed companies today tend to be older and larger than in previous years. As the average age of IPOs rose from 8 to 11 years, there would have been a temporary slowdown in new issues as companies grew older. However, after 16 years that slowdown is behind us and does not explain the weak numbers of the last decade or more.

Two reasons often cited to explain the missing IPOs are the regulatory burden imposed on public companies and fears of activist investors. The law that matters for regulation is more likely Sarbanes–Oxley than Dodd–Frank.

Figure 4. US Initial Public Offerings Annually



Source: Jay R. Ritter, University of Florida, August 2017.

¹⁰²Jay R. Ritter, “Initial Public Offerings: Updated Statistics,” August 2017, <https://site.warrington.ufl.edu/ritter/ipo-data/>.

The latter focuses on the financial sector, whereas Sarbanes–Oxley broadly targets public companies with regulations on reporting and certifying financial results.

Founders taking their company public do not want to risk losing control of their company. Whether the activists are more of a threat to IPOs in particular or to other companies isn't clear. What is clear is that many managers, especially in the technology sector, are worried enough to create corporate structures that insulate the board of directors and management from shareholders. These structures often include multiple classes of stock and differential voting rights that favor the founders or incumbent management. A related argument is that private companies can plan for long-term results, whereas a public company may be judged only on last quarter's earnings.

The potential returns to investing in IPOs changed as well. Of course, IPOs are still high-risk, high-return opportunities. However, today's companies are older and larger when they go public. Therefore, opportunities for outsized price jumps on the first day of trading and extremely large long-run returns appear harder to find. Consider three famous cases for investors who held on to the stock: Amazon went public when it was three years old, raised \$625 million and the return was 565 times investor investment; Google was six years old at its offering and investors made 20 times; Facebook was eight years old, and investors made almost four times their money.¹⁰³

Encouraged by the strong stock market in 2017, IPOs experienced their best year since the end of the financial crisis. Activity in the United States saw a resurgence, and China set a record for deals. American IPOs raised twice as much capital in 2017 than in 2016. Whether this pace can be sustained remains to be seen.¹⁰⁴

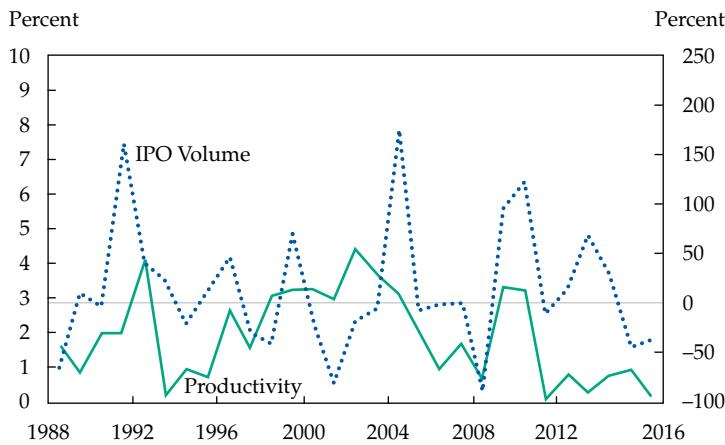
A recovery in IPOs would be a positive sign for productivity. As **Figure 5** shows, IPOs and increases in productivity tend to move together. The co-movement of IPOs and productivity reflects a willingness on the part of businesses and entrepreneurs to invest. Although IPOs are not the only source of capital to support productivity gains, the data in Figure 5 suggest that IPOs are an important factor.

Investors are turning to other ways to participate in the growth of startups and new companies. The private market is a growing source of capital.

¹⁰³Michael Mauboussin, Dan Callahan, and Darius Majd, "The Incredible Shrinking Universe of Stocks" (Credit Suisse, 22 March 2017).

¹⁰⁴Nicole Bullock and Robert Smith, "Global Number of IPOs Highest since Financial Crisis," *Financial Times*, 27 December 2017.

Figure 5. Year-over-Year Change in Productivity and IPO Volume



Sources: US Bureau of Labor Statistics Nonfarm Business Labor Productivity; Jay R. Ritter, “Initial Public Offers Statistics,” University of Florida, August 2017.

Where Capital Comes From Now

Private market funding of startup companies is large and growing. Unicorns—that is, private companies with valuations beyond a billion dollars—are an expanding list. The best-known unicorn is Uber, valued at \$60 billion. Other well-known US names include AirBnB, WeWork, and DropBox. Each is valued at over \$10 billion. The United States has the largest concentration of unicorns but is far from being the only country with them. China boasts about 60, and India and the United Kingdom have several each.¹⁰⁵

Even if all the unicorns were to become public, the dearth of IPOs and missing listings will not be solved. Companies are still going public, albeit more slowly and later. However, as big or rich as the unicorns are, there are not enough of them to make up for the 4,000 to 5,000 missing listings or to reduce concentration among many industries.

In late 2017 there were 108 US unicorns valued at \$382 billion. The United States represents about half of the global total of 216 unicorns, worth a combined \$757 billion. China ranks second, with 58 companies worth \$256 billion; the United Kingdom is third among countries, with 12 entries worth \$22 billion; and India is fourth, with 10 unicorns worth \$35 billion. The oldest of these became unicorns in 2009.¹⁰⁶

¹⁰⁵<https://techcrunch.com/unicorn-leaderboard>, <https://www.crunchbase.com/>.

¹⁰⁶<https://www.cbinsights.com/research-unicorn-companies>.

Commentators and analysts celebrate unicorns as evidence that the US economy is vital, growing and creating new companies and products all the time. Unicorns may in fact be doing all this. At the same time, investment opportunities are being siphoned from the public markets as many former unicorns never go public but instead disappear into the shrinking list of public companies. Others may become IPOs and be acquired soon thereafter.

Some 110 companies initially made the unicorn list and then dropped off. Of these, 49 were acquired and 61 went public. Those that went public include household names like Facebook, Tesla, and Pandora. However, some of the IPOs were subsequently acquired, including BATS, LinkedIn, and Zappos. Still others like WhatsApp or Instagram were acquired before they became public.

Nonetheless, a public offering can provide real benefits to companies and their shareholders—namely, much greater liquidity for existing shareholders, a mechanism to offer stock-based compensation to employees, more widely accepted valuations compared to wishful thinking on the private markets, and most of all, fresh capital.

Capital in New Shapes¹⁰⁷

Hidden behind the IPO drought and the rise of unicorns is the changing nature of capital itself. Capital today, like the markets, isn't exactly what it was 20 years ago.

In the official National Income and Product Accounts, Nonresidential Fixed Investment (a component of GDP) was traditionally defined as “structures” and “plant and equipment.” In 2013 a new category, intellectual capital, was added that includes just about everything else—for example, software, patents, video games, and entertainment. The GDP data on capital investment flows show the growing importance of intellectual capital and the declining significance of structures, property, and plant and equipment.

The share of intellectual capital in annual capital investment doubled from 17%, or one-sixth, in 1985 to 33% in 2016. Intellectual capital grew by 5.6% annually in real (inflation-adjusted) terms. This compares to a growth rate of 4.5% for equipment and a decline of 0.4% annually for structures over the 1985–2016 period. The trend shows no sign of reversing.¹⁰⁸

These figures reflect capital investments on a company’s balance sheet, but they do not tell the whole story. Today capital investment is also in the income statement. Not all intellectual capital is purchased in the market; it

¹⁰⁷Jonathan Haskel and Stian Westlake, *Capitalism without Capital: The Rise of the Intangible Economy* (Princeton, NJ: Princeton University Press, 2017).

¹⁰⁸US Bureau of Economic Analysis, National Income and Product Accounts.

is also homegrown, created with human capital and paid for through wages, salaries, and other compensation. Other forms of intellectual capital that are licensed, rented, or paid for through royalties also appear on the income statement. Further, many services of traditional capital are easier to rent than to buy. Companies may use cloud-based computing services instead of buying or building data centers. Others develop business models where someone else buys the capital equipment—Uber doesn't own all its cars, nor does Google own all the websites its employees search through for all of us. The nature of capital today is intrinsically different from the past, and companies may not, or cannot, control all the capital essential to their business.

Looking over today's market, one can see enormous changes: There are half as many stocks as 20 years ago, capital is leaving the market through buybacks, and the market is being rearranged by mergers and not being replenished by IPOs. Much of today's capital would not have been counted, and might indeed not have existed 10 or 20 years ago. One result of these shifts is that the public stock market's role as the primary market for raising capital is diminished.

This is not the end of public equity markets. The primary public market—where new capital is raised—may in fact have shrunk or been partially replaced by a growing private primary market for capital. However, all primary markets depend on the price discovery and valuations derived from trading in the public equity market. Moreover, only a minority of investors can access the private equity market, so the stock market remains the focus of savings and equity investment in the economy.

Why Is Business Investment So Low in the United States?

Thomas Philippon

Professor of Finance, New York University Stern School of Business

Introduction

Corporate investment in the United States, when stated as a percentage of profits, has dropped in recent years and is now at lows not seen in several decades. (By “investment,” I mean real investment, for example, in factories, in plant and equipment, or in intellectual property.) This article provides evidence of this trend and suggests that its primary cause is an increase in industry concentration and the related decline in competition. An uptick in regulation appears to be another factor in this predominantly US phenomenon as European companies continue to invest at earlier levels even though Europe has experienced its own financial crisis.

Corporate Investment Is Low Relative to Fundamentals

To answer this question, we looked at various measures of corporate profits and funding costs, each of which shows that investment is low relative to the amount of money firms are actually making. **Figure 1** shows net investment (gross investment less depreciation) as a percentage of profits (net operating surplus).¹⁰⁹

Historically, net investment has averaged 20% of net operating surplus; that is, for every dollar earned, firms plowed 20 cents back into investment. In recent years, this ratio has dropped to 10 cents on the dollar.

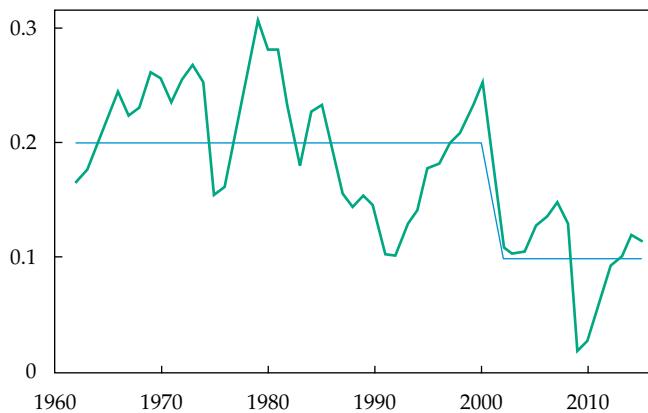
In terms of whether or how much to invest, it is not just current profits but expected future profits that matter (taking into account risk premia and other factors). **Figure 2** attempts to explain this principle using Tobin’s Q , which measures the capitalized value of future profits and which historically has done a good job of explaining corporate investment.

In Figure 2, the dotted line shows net investment predicted by Tobin’s Q in the United States, over time. Tobin’s Q is defined as the market value of assets divided by the replacement costs of these assets. When Q is above 1, firms should invest and assets should grow. The green line shows the

¹⁰⁹The growth rate of the capital stock is typically 3% per year.

Figure 1. Corporate Real Investment Is Low Relative to Profits

Net Investment/Net Operating Surplus

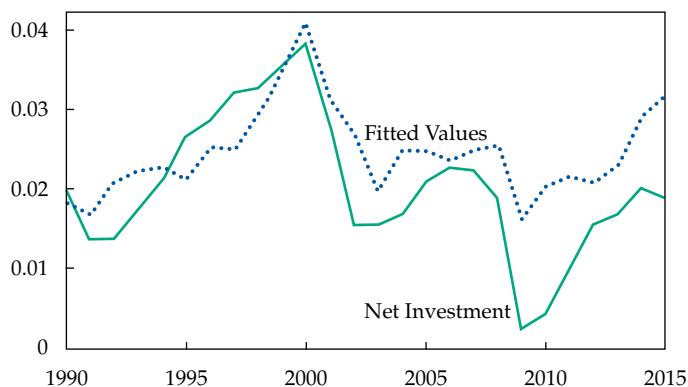


Notes: Annual data. Net investment/net operating surplus for non-financial business sector.

Source: US Flow of Funds Accounts.

Figure 2. Corporate Real Investment Is Low Relative to Tobin's Q

NI/K



Notes: Annual data. The green line shows the observed net investment rate (NI/K) and the dotted line shows the net investment rate predicted by Tobin's Q. Net investment and Q are calculated for the non-financial business sector.

Sources: US Flow of Funds accounts. See also Gutierrez and Philippon (2017).

investment rate, reflecting both the corporate and noncorporate business sectors. Because the capital stock is, effectively, the cumulative sum of business investment in each period, we can estimate from this figure that the capital stock was 10% smaller by 2015 than what would be predicted given the level of Tobin's Q . This shortfall, which is significant and not the result of deteriorating fundamentals, has long-run negative effects on future wages and productivity.

Industry Concentration and Lack of Competition Are the Primary Drivers

Instead of directly asking *why* this is the case, let's first start by asking the opposite question: What are some plausible reasons for low corporate investment that do not happen to be the case? First, low corporate investment is not the result of financial constraints (at either the firm or the industry level). Pessimism about the future is another nonissue, since investor expectations are taken into account in capitalized future asset values.

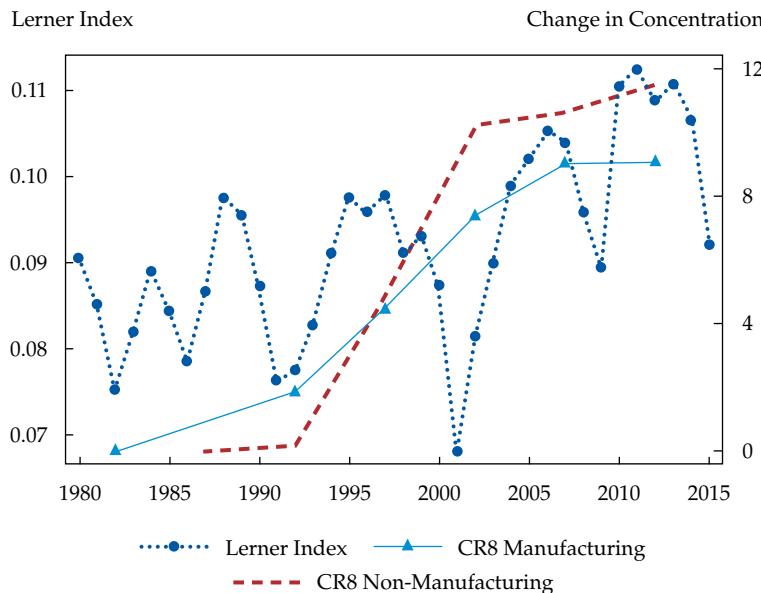
Further, low corporate investment cannot be explained by high uncertainty or low productivity growth; if that were the case, it would have been captured using Tobin's Q . To be clear, productivity growth may indeed have slowed, but that factor does not explain the lack of investment relative to where it should be, given the fundamentals. Productivity is one of these fundamentals—if it happens to be low, Tobin's Q could be low—but that effect is not what is driving the data.

What is it, then? It turns out that the increased concentration of US industries (and the related lack of competition) is the primary driver of lower corporate investment. According to this decreasing domestic competition (DDC) hypothesis,¹¹⁰ when firms operate in an environment that is not competitive, they become highly profitable—which explains their high Tobin's Q —but they no longer have an incentive to invest (see Philippon 2019 for a complete discussion). It is important to emphasize the term *domestic* because there are subsectors in the United States that face increasing foreign competition, although that is not broadly the case.

Figure 3 shows the relationship between profits and industry concentration since 1980 for both manufacturing and non-manufacturing (service) sectors. The dashed (non-manufacturing) and thin solid (manufacturing) lines are measures of concentration for the universe of all firms—there is no bias in the sample as every firm is represented. Both sectors show a sharp increase in concentration. The dotted, more volatile line is the Lerner index, a measure of

¹¹⁰See Gutierrez and Philippon (2017).

Figure 3. Profits and Concentration



Notes: The Lerner Index from Compustat is defined as operating income before depreciation minus depreciation divided by sales. The eight-firm CR from Economic Census is defined as the market share (by sales) of the eight largest firms in each industry. Data before 1992 are based on SIC codes. Data after 1997 are based on NAICS codes. Data for manufacturing are reported at NAICS Level 6 (SIC 4) because the data are only available at that granularity in 1992. Data for non-manufacturing are based on NAICS Level 3 segments (SIC 2).

Sources: The Lerner Index from Compustat. CR8 from Economic Census. See also Gutierrez and Philippon (2017).

the profit margin that firms are making.¹¹¹ Starting in 2000, profits began to trend upward as these sectors became more concentrated, which is consistent with the DDC hypothesis.

As another important piece of evidence of the effects of the increase in concentration, Gutierrez and Philippon (2017) show that the entire investment gap is coming from industries that are becoming more concentrated. A sharp increase in the Herfindahl index (more concentration) of the top 10 most concentrated industries occurred at roughly the same time as investment as a percentage of profits for these industries began to decline. As a result, their

¹¹¹The Lerner index, set forth in 1934 by economist Abba Lerner, describes the extent to which a firm has monopoly power. The numerical value of the index is $(P - MC)/MC$ where P is the price that a firm sets for a product and MC is the firm's marginal cost for that product. The index ranges from zero to 1, with higher numbers representing greater monopoly power.

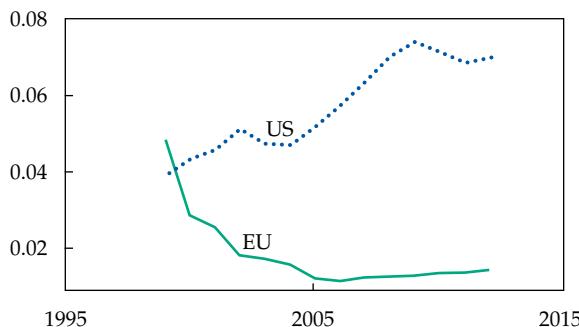
capital stock is now 20% below where it was in 2000 relative to Tobin's Q. If we look instead at industries that have not become more concentrated, we do not find a gap between what they invest and what is predicted by Tobin's Q.

A US Phenomenon

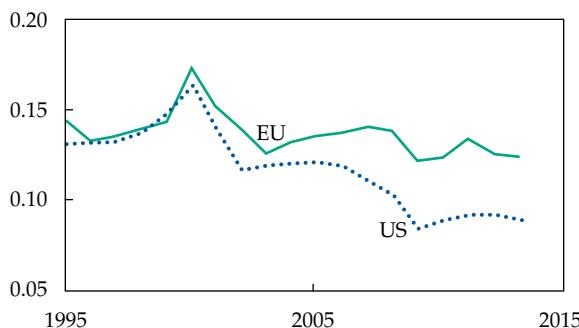
The evidence suggests that this is not a global trend but rather a phenomenon almost exclusive to the United States. **Figure 4** compares the United States

Figure 4. European Union vs. the United States: Concentration and Investment in Five Industries where Concentration Increased in the United States

A. Herfindahl



B. Investment Rate (I/K)



Notes: The figure is based on the top five concentrating industries in the United States: information telecom, arts and recreation, wholesale and retail trade, other services, and information publishing (which includes software). It excludes textiles because of import competition. Panel A plots the weighted average Herfindahl across these industries, weighted by sale. For the European Union, each industry's Herfindahl is the weighted average Herfindahl across countries. Panel B plots the weighted average investment rate, weighted by the capital stock.

Source: Bureau of Economic Analysis. See Gutierrez and Philippon (2017).

with Europe in terms of both concentration (Panel A) and investment (Panel B) for the five most rapidly concentrating industries in the United States.¹¹² Panel A shows that whereas concentration has gone up in the United States, it has been going down in Europe. In Panel B, it is clear that European companies in these same industries, despite their more recent financial crisis, have invested at least as much, and often more, than their US counterparts.

To expand on this point, let's drill down further and look at the telecom industry in both the United States and Europe. It is currently twice as expensive to buy and operate a cell phone in the United States as it is in France, a fact that is not widely known. This is a change from 1999 when I first came to the United States. At the time, I was struck by three things: (1) A laptop in the United States was 40% cheaper than in France, (2) local calls were free (which was not the case in France at the time) and cell phones were significantly less expensive, and (3) airplane tickets were so much cheaper in the United States that you could fly almost anywhere at any time.

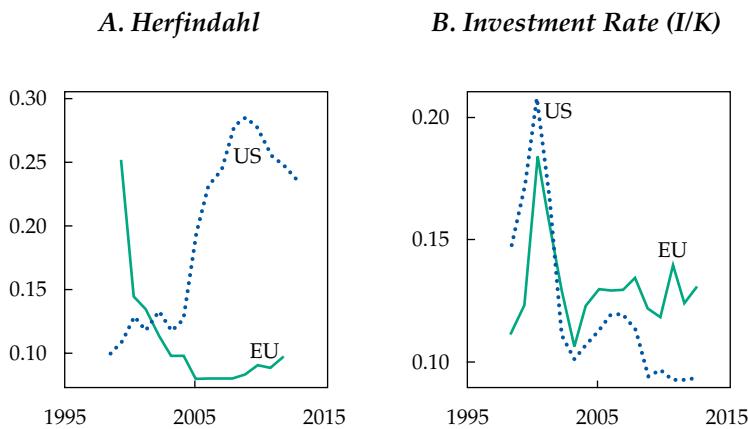
Every single one of these facts is now reversed. It costs half as much to own a cell phone in France than it does in New York. Airline tickets are much cheaper in Europe. Laptops are the same price. This happened over the past 20 years.

Figure 5 shows concentration and investment for the United States and Europe for the telecom industry. Concentration has gone up in the United States (with investment crashing) and has come down in Europe (where the investment rate is higher).

The left panel of **Figure 6** shows concentration in the airline industry for Europe and the United States. The right panel shows the price-cost margin, which is the price minus the marginal cost as percentage of the price $((P - MC)/P)$ and is commonly used as a measure of competition in an industry. As evident, after 2005 the airline industry in the United States got very concentrated, and this is reflected in an increased price-cost margin. Meanwhile, in Europe we do not see any trend. According to *The Economist* (2017), the airline industry is one of the top five most concentrating industries in the United States, with the top four airlines in the country having an 80% market share, up materially from a decade or two ago; this number is 40% in Europe. And 80% doesn't tell the whole story, as US airlines have structured the markets such that on any given route, only two airlines will typically compete, and in some smaller markets only one is present.

¹¹²The same industry classifications (SEIC codes) were used in the United States and Europe.

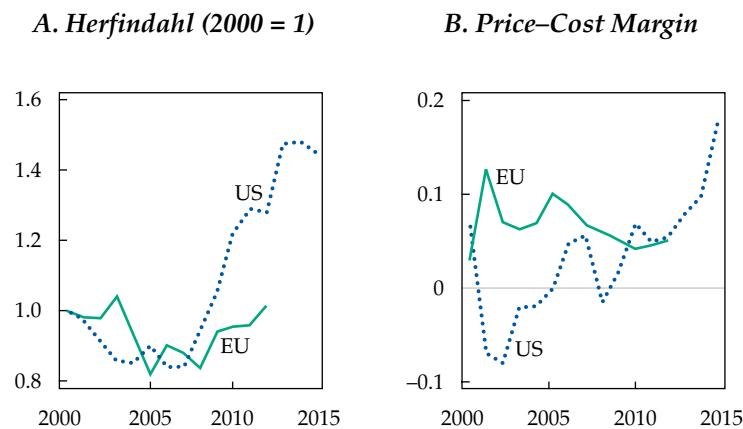
Figure 5. Changes in Industrial Concentration Indices, European Union vs. the United States: The Case of Telecom



Notes: Comparison of the evolution of the Herfindahl and investment rate in the telecom industry for the United States and Europe.

Sources: European series are based on the weighted average across major EU economies, series is based on CompNET. The US series is based on Compustat. See Gutierrez and Philippon (2017) and Faccio and Zingales (2017).

Figure 6. Changes in Industrial Concentration Indices, European Union vs. the United States: The Case of Air Transportation



Notes: Comparison of the evolution of the Herfindahl and price-cost margin in the US and EU air transportation industries. The price-cost margin is the price minus the marginal cost as percentage of the price $((P - MC)/P)$.

Sources: Compustat for the US data and CompNet for Europe. See Gutierrez and Philippon (2017).

Because airlines have split the routes among themselves in this way, they make two to three times as much money per passenger mile as do European airlines. In addition, the United States government forbids foreign airlines from competing in the United States; they are not allowed to fly US domestic routes. This is in contrast to Europe, where such access and competition are encouraged.

A Role for Product and Market Regulation

We then looked to see if industry regulation could explain some of the trend toward lower investment in the United States. When I began to study economics in earnest in the 1990s, the periodic OECD¹¹³ reports were as you would expect, e.g., France's labor and product markets were too regulated, particularly relative to the United States. It was the same report every four years, and it was getting dull; I stopped reading the reports around 2000. Flash forward to today and it is a different story. In 1997, every major European country (Germany, Italy, France, etc.) had more product market regulations than the United States. Over the next 20 years, however, this situation began to change such that in 2013, every one of those European countries had less product and market regulation than the United States.¹¹⁴

Threat of Competition Drives Investment

When looking at the life-cycle dynamics of an established industry, you see leaders and laggards. When new firms come in, they are typically better than and eventually kill off the laggards. They tend to invest more and are more productive than the established players. At the industry level, the direct effect of new competitors is that prices tend to drop and overall investment increases.

There is also an indirect effect on the leaders in the industry. Once-quiet industry leaders tend to react by investing more in both intangible and tangible assets. Using Compustat data, Gutierrez and Philippon (2017) show that the entire gap in business investment and the accumulated capital stock is driven by underinvestment by the leaders in concentrating industries.

Given that the measures of competition are imperfect, in some cases you could get misleading correlations. That is, concentration is not necessarily causing lower investment; it could be that the same factor is driving both concentration and investment. Here is a simple and realistic example: In a

¹¹³Organization for Economic Co-Operation and Development.

¹¹⁴See OECD (2013) report and Döttling, Gutierrez, and Philippon (2017). See Philippon (2019) for a discussion of the political economy of regulation.

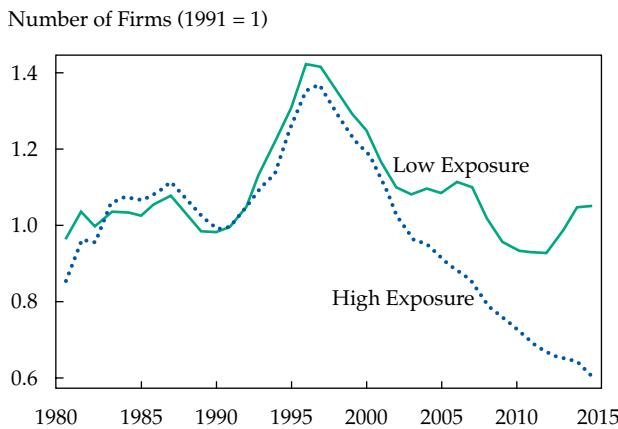
declining industry you are likely to see both mergers (to control or cut costs) and less investment. Concentration could be rising if the industry has been in decline for a long time, but you wouldn't say that concentration, in and of itself, is causing the lack of investment. Rather, it is the decline of the industry that causes both the decrease in investment and the increase in concentration.

To test if a causal relationship exists, we used a handful of techniques, starting first with the exception that proves the rule. As mentioned, most US industries have become less competitive, but that is not the case for industries exposed to foreign competitors.

We can use China's entry into the World Trade Organization (WTO) as a specific marker of increased competition and can measure the effect of that event on specific US industries.¹¹⁵ When China started to compete with the United States, it killed the firms that were not very competitive in the first place, the laggards. **Figure 7** shows two groups of industries, those with low import exposure to China and those with high import exposure.

Pre-WTO, these two groups track in terms of the number of firms, including during the tech bubble and bust of the late 1990s. The break occurred in 2000 when China entered the WTO. From there on, the number

Figure 7. Number of US Firms, by Exposure to China (Compustat)



Notes: Annual data of US incorporated firms in manufacturing industries only. Industries are assigned to exposure based on median 1991–2011 exposure (1991 = 1).

Source: Firm data from Compustat; import data from UN Comtrade.

¹¹⁵Chinese access to US markets is a discrete event allowing us to measure the direct effect of increased competition on an industry not affected by other forces such as changing technology or demand dynamics.

of firms fell drastically in the industries exposed to China as part of a cleansing effect; the low-productivity firms were kicked out. What is even more interesting is how the leaders in these industries reacted.

The leading US firms were highly productive and did not have much domestic competition. All of a sudden, with the entry of China into their markets, they began to feel more competitive pressures. As a response, Gutierrez and Philippon (2017) show, leaders of industries exposed to the China shock invested substantially more than laggards to counter the threat posed by the new entrants. In contrast, in industries that were not exposed, leaders and laggards continued investing at the same rate.

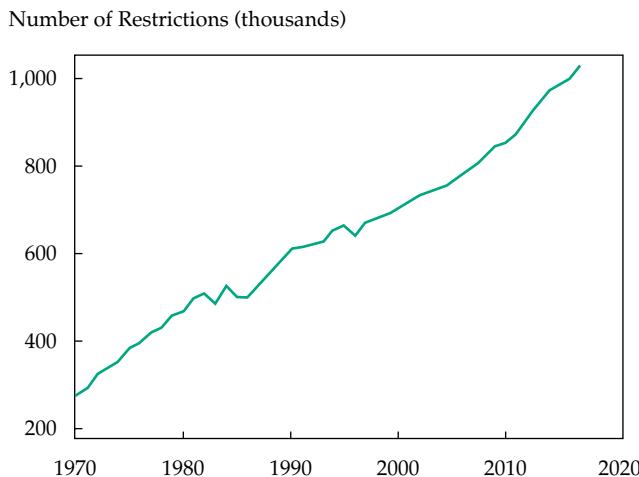
This additional investment, when you drill down, was substantially directed toward intangibles as the leaders tried to innovate their way out of the threat posed by foreign competition. The evidence suggests that firms threatened by competition invest and innovate more. The flip side is that if they are not threatened, they tend to increase dividends and/or share buy-backs rather than invest or innovate.

What Changed in 2000?

It appears from the data that most of the changes in concentration, competition, and ultimately investment in the United States occurred after the year 2000. Although this is a bit more difficult to tease out, the comparison with Europe suggests that there should be a role for antitrust regulation. The example of the telecom industry is obvious; that regulators in Europe have been much more aggressive in promoting competition than the US airlines is another. Furthermore, at least one or two mergers in the United States were clearly anticompetitive. In Europe, on the other hand, airline regulators have been aggressive in not allowing mergers that would thwart competition.

Being careful not to pick up the effects of decline in an industry where concentration is naturally occurring as is the resulting slowing of investment, Gutierrez and Philippon (2017) looked at what happens to investment after a large merger. The good thing with mergers is that they are very lumpy and, in a declining industry, tend to happen all at once. This fact provides a natural experiment that allows us to tease out the causal impact of competition. We find that overall investment declines leading up to the merger, then plummets after the event. This turns out to be a significant contributor to the decline in corporate investment since 2000, given that large mergers have been a key feature in the United States over the past 15 years.

The other factor is regulation, which has increased in the United States since 2000. The OECD data discussed earlier and the Regdata Regulation Index (**Figure 8**) both show this. With the OECD data, Gutierrez and

Figure 8. Regulatory Accumulation since 1970

Notes: This *y*-axis is the total number of restrictions in the United States. Restrictions, which use words like *shall* and *must* that indicate prohibited or required activity, are published annually in the Code of Federal Regulations (CFR).

Source: Regdata 3.1, available at <https://quantgov.org/regdata/>.

Philippon (2017) find that industries where regulation increases the most end up with fewer firms, higher profits, and less investment.

Conclusion

Most industries in the United States have become more concentrated and less competitive. This reduction in competition explains close to two-thirds of the investment gap we have seen since 2000. When competition is introduced (e.g., China's entry into the WTO), industry leaders respond with investment in intangibles, to a substantial degree, but with capital expenditures as well.

Banking is an excellent example with which to conclude. There is a general belief that banking in the United States is more competitive than in Europe, which is dominated by large banks that dictate what they want in terms of regulation. But this state of affairs is changing, and European banking could easily become more competitive than US banking within five years. The first change is in the way regulations are enforced in Europe, as evidenced by the Single Supervisory Mechanism (SSM) in Frankfurt.

Another striking development is related to financial technology (fintech). Many startup firms have a business model that aggregates all of a consumer's financial data in one place and offers low-fee products and advice. To do this,

they need access to the customer's information (banking history, retirement plans, investments, etc.). Banks are typically opposed to these evolutions because they are worried about losing their relationship with their clients. In Europe, it was determined that consumers are the owners of their data and therefore have the right to say who has access to it. European banks will therefore be required to build an application-programming interface (API) so that consumers can securely login and transfer data. In Washington, however, the banking lobby prevailed. Thus US consumers do not effectively own their data and competition remains limited. Even in banking, then, the United States could lose its competitive edge if nothing is done.

References

Döttling, R., G. Gutierrez, and T. Philippon. 2017. "Is There an Investment Gap in Advanced Economies? If So, Why?" Paper presented at the European Central Bank Forum on Central Banking Conference on Investment and Growth in Advanced Economies, Sintra, Portugal, 26–28 June.

The Economist. 2017. "A Lack of Competition Explains the Flaws in American Aviation." <https://www.economist.com/leaders/2017/04/22/a-lack-of-competition-explains-the-flaws-in-american-aviation>.

Faccio, M., and L. Zingales. 2017. "Political Determinants of Competition in the Mobile Telecommunication Industry." NBER Working Paper 23041.

Gutierrez, G., and T. Philippon. 2017. "Decreasing Competition and Investment in the U.S." NBER Working Paper 23583.

———. 2018. "How EU Markets Became More Competitive than US Markets: A Study of Institutional Drift." NBER Working Paper 24700.

OECD. 2013. Product Market Regulation Database. www.oecd.org/economy/pmr.

Peters, R. H., and L. A. Taylor. 2017. "Intangible Capital and the Investment-q Relation." *Journal of Financial Economics* 123 (2): 251–72.

Philippon, T. 2019. *The Great Reversal: How America Gave Up on Free Markets*. Cambridge, MA: Harvard University Press.

Pierce, J. R., and P. K. Schott. 2012. "A Concordance between Ten-Digit US Harmonized System Codes and SIC/NAICS Product Classes and Industries." *Journal of Economic and Social Measurement* 37 (1–2): 61–96.

Regdata 3.1 available at <https://quantgov.org/regdata/>.

Solutions and Proposals

Beyond Blue, Part 3: The Power of Infostructure

Walter Russell Mead

Ravenel B. Curry III Distinguished Fellow in Strategy and Statesmanship, Hudson Institute, and the James Clarke Chace Professor of Foreign Affairs and Humanities, Bard College

The quest for a new social model has to start with economics. America could survive without growing prosperity and rising standards of living, but it would not flourish—and it would not be living up to its potential to create a better life not only for Americans but for people all over the world. Green dreamers and communitarians disagree, often eloquently but always futilely; the drive for economic prosperity is deeply planted in American politics and society. When the economy isn't performing well, politicians lose their jobs while the public looks for alternative ideas.

The quest for economic prosperity helped make the blue social model, which I described in Mead (2012), and the failure of that model to deliver continuing prosperity in contemporary conditions is both a symptom of and a leading reason for its decline. The mass prosperity of Fordism depended on economic conditions that no longer hold. Highly paid manufacturing jobs and many clerical and middle management positions are disappearing. Some jobs are outsourced to cheaper foreign countries; others disappear as automation increases productivity and decreases the number of people needed to accomplish various tasks. The stable oligopolies and monopolies that once dominated the American economy have disappeared in the face of heightened international competition and accelerated technological change. As formerly large and stably profitable corporations had to scramble to survive, they no longer could afford lifetime employment, friendly relations with strong unions, and generous health care and retirement benefits. Economically, the decline of the blue social model presents Americans with some urgent questions: How can we generate a rising standard of living for our citizens in a post-Fordist world? If manufacturing and stable oligopolies won't underpin lifetime employment and rising wages for new generations of Americans, what are we going to do instead? To start answering these questions we have to think about whether the United States enjoys any comparative advantages

Reprinted from *The American Interest*, 5 February 2012, with the permission of Walter Russell Mead.

or we can develop some that, wisely deployed, could provide us with the kind of rising living standards that we have enjoyed in the past? It's clear that many such advantages exist and some are even becoming more prominent and useful. In the energy sector, homegrown shale gas and shale oil plus large new discoveries in the Western Hemisphere suggest that 21st century America will enjoy secure access to an abundant and varied mix of fuel from nearby and friendly neighbors. Blessed by a favorable climate and rich soil, our agricultural productivity will continue to astound. Our geographical location, which gives us access to both of the world's greatest ocean trading basins while protecting us against invasion by other great powers, remains a tremendous advantage. The language we speak continues to develop as the global *lingua franca*, giving every American citizen and American business a significant boost.

These are advantages we have enjoyed for centuries and they are likely to loom larger in the 21st century even than in the recent past. We have grown proficient at using and enhancing these advantages and we need to continue to do so moving forward. Just based on these factors, fears that we are sinking back to some kind of dystopian, third world future are, I think, overblown. But our greatest advantages are cultural and political, and it is to these that we must ultimately look for the prosperity we want in the decades to come. I've written about these advantages at some length in *God and Gold: Britain, America and the Making of the Modern World*, and I won't repeat those arguments here. The upshot is that, among all the countries in the world, the United States and a handful of English-speaking societies have some unique strengths. We do better at combining rapid innovation with social and political stability than others do. Since 1688, the English-speaking world has seen only two successful violent revolutions, and even the American Revolution was largely non-violent with royalist governors peacefully replaced until the British army arrived to reconquer the colonies. (The other successful violent revolution was the secession of Ireland from the United Kingdom in 1919.) The United Kingdom, the United States, Canada, Australia, New Zealand, and, since independence, Ireland have had remarkably stable political histories and show great continuities in culture even as they have remained at the forefront of economic and technological change. Despite a history of booms and busts from the South Sea Bubble to the 2008 meltdown, their financial histories have also been more stable than those of other countries; since the establishment of the Bank of England in the late 17th century, the political stability of the English-speaking world has contributed to economic stability. Investors have been able to buy long-term bonds in the English-speaking world with confidence since the 18th century. No other significant global

economic or political centers have this kind of track record, though the Dutch come close. (Ugly budget pictures today in Ireland, the United States, and the United Kingdom tell us that we should not take these economic advantages for granted. Fortunately, in all three countries there seems to be some real political will to turn things around.) The English-speaking world has been at the heart of the process of modernization with all the upheaval and uncertainty this process brings in train, but through it all these countries have somehow remained more stable and reliable than others. We change faster than others do in response to new opportunities and new technological possibilities—but we don't lose our balance in the process.

Joseph Addison described the behavior of Winston Churchill's famous ancestor the Duke of Marlborough at the battle of Blenheim in a way that captures this ideal of calm in the midst of frenzied activity:

The dreadful burst of cannon rend the skies,
And all the thunder of the battle rise.
'Twas then great Marlborough's mighty soul was prov'd,
That, in the shock of charging hosts unmov'd,

Amidst confusion, horror, and despair,
Examin'd all the dreadful scenes of war:
In peaceful thought the field of death survey'd,
To fainting squadrons sent the timely aid,
Inspir'd repuls'd battalions to engage,
And taught the doubtful battle where to rage.
So when an angel by divine command
With rising tempests shaks a guilty land,
Such as of late o'er pale Britannia past,

Calm and serene he drives the furious blast;
And, pleas'd th' Almighty's orders to perform,
Rides in the whirlwind, and directs the storm.

This angel in the whirlwind quality means more now than ever before. The 21st century is shaping up to be an age of upheaval; change is coming at us from so many directions and at such a pace that cultures and countries around the world are being shaken to their foundations. Those who can keep their calm and balance in the midst of the whirlwind have a serious advantage—and they should use it for all it is worth.

America is good at change. We absorb immigrants better than most. We like new things and like to try them out. We have an optimistic streak in our nature; we believe that change is basically good and that being open to new things will make us happier and better off. Our religious sensibility is future

oriented and believes that God is working through the chaos and uncertainties of life. Our national religious tradition is profoundly influenced by the dynamic vision of a God who calls humanity into an unknown future. While the religious cultures of some parts of the world look back to a real or imagined utopia in the far distant past, or instruct the faithful to resist change and cling to the ancient ways, American religion tends to see the hand of God behind the winds of change. We pursue God into the future, rather than hunting for him in the far-distant past.

America's critical comparative advantage in the 21st century will be its ability to respond quickly to change: to recognize and exploit new opportunities faster than others, to retool its core institutions and practices to fit the emerging shape of the new world, and to do all that while retaining its political and social equipoise: to ride the whirlwind and direct the storm. We were the first to build the blue social model and we can be the first to get to the next stage and reap the enormous rewards that come from reaching a more productive and efficient form of social organization before the competition.

No doubt many new opportunities will emerge during the 21st century; nanotechnology and biology look to provide revolutions at least as profound as anything to be found in the information technology (IT) revolution now transforming the world. But until these new developments come more fully on line, the biggest opportunity and the greatest challenge that faces all the leading economies will be to harness the full power of IT to social needs.

All about Friction

In particular, this means competing with other countries not by the cheapness of our wages or the laxity of our environmental regulations, but by building on our ability to increase the productivity of both capital and labor through the power of IT to reduce the cost of friction in society. The first challenge of the 21st century will be the race to build infostructure—a mix of hardware, bandwidth, software, and government and corporate practices that deliver the greatest possible benefits of IT in ways that dramatically reduce costs and delays throughout the economy. A lot of this will be about friction.

Readers of Clausewitz know that friction was his word for the inexorable entropic pressure that disrupts any plan of campaign. Bad weather, diseases in the camp, miscommunication of orders, the absence of good information, the hazards of a night march: All these could be examples of the kind of friction with which every commander must struggle.

In civilian life as well as military affairs, friction has often been caused by distance: Moving people, goods, and information through space costs money and time. We will continue to build and repair physical infrastructure in the

future, but the cutting edge of society's effort to reduce frictional costs in time and money will lie in the development of "infostructure" rather than infrastructure: the mix of institutions, practices, hardware, software, and bandwidth that allows the inherent productivity of information technology the fullest possible scope to transform the way we do things.

Take the legal system. The legal system arose in order to solve disputes; it consumes a great deal of time and money and is itself one of the great sources of friction in our society, but it is much better than anarchy would be. Yet as society becomes more complex, we need more legal service and not less, and the question of how to make the legal system fast, efficient, and cheap grows more important all the time.

Let us imagine how the legal system might work some years down the road with the right kind of infostructure. Some changes are already under way. Routine matters like basic simple wills, taxes, and incorporations can be handled using cheap commercial software. Clearly, this software is going to get better and smarter, and more and more matters can be handled by people working on their own. Governments should be working to facilitate this shift, changing laws where necessary and advisable. Perhaps one day most routine marriages and divorces, residential real estate transactions, and many other activities that historically required lawyers can be done by individuals using software or at most with paralegals. More and more of this will move into the "cloud"; less and less time and money will be chewed up by routine legal affairs.

That is all first-stage reform. Deeper change and greater savings will come as more sophisticated software and new institutions and practices that take full advantage of these capabilities transform the inner workings of the legal system. It's impossible to foresee how this would work out in practice—any more than someone in 1895 could predict how the modern chemical industry would develop over time. But develop it will. One possible way things could go: Every court decision ever given, every brief ever filed, and every journal article ever written will someday be available in a single database; intelligent software could sift the facts of any given case and the arguments of both sides against this material to make some predictions about how a given case might come out. Given access to the records and decisions of the judge appointed to hear a particular case, it would be possible to see how this particular judge would be likely to view this particular case with a fair degree of probability.

That information would be available to both parties in the case, making a fast settlement a much more likely outcome. Many fewer cases might go to trial; decisions would in any case come much more quickly.

Smart policy would speed this transition. Investing in the infostructure that moves the state's legal system as far into the future as quickly as possible would attract new investment and jobs and enhance the profitability of business already in the state.

Research into ways to use IT to accelerate the speed of legal proceedings while reducing their cost is the kind of thing government should support. Land grant colleges supported agricultural research in the age of the family farm; we developed a host of ways to support research into improved industrial and factory productivity in the age of Fordism. Now it is time to shift to the next stage, for universities and government to move into the kind of research that will transform our social infostructure.

While infrastructure will remain important to the smooth functioning of society, infostructure will likely surpass it. Governor Jerry Brown's successors may think less about building \$100 billion high-speed rail programs than about radical restructuring of the way California's government works so that a much smaller and cheaper government is able to do more and do it much quicker. A superfast and supercheap legal system would dramatically enhance the productivity of capital invested in the state. Government would work better and faster and cost less; taxes would fall even as services improved. (There is no real reason why many government services can't improve the way personal computers and software have done: Each year they get cheaper and each year they do more.)

The war on friction is one of humanity's oldest challenges, and it inevitably becomes more important as society becomes more complicated and more interdependent. A small village can handle everything quickly; a modern megalopolis must have a much better developed government. Good infostructure can dramatically reduce the costs and complexities of modern governance.

America's Secret Weapon

In all societies vested interests will resist efforts to reduce friction. The mule drivers didn't like the canals, and the canals hated the railroads. Pettifogging lawyers will hate the transformation of their profession, and every vested interest in the country will be touched one way or another by the kinds of transformations of our core social institutions that will take place as we build out the infostructure of the 21st century.

This is where America's advantages count. That process can and should happen faster here than in other places. We have 50 states that compete with one another and the federal government to do things faster, quicker, better. Best practices and new ideas spread. Bureaucrats are less powerful here, and

state employees are less well placed to block changes they don't like. The public is readier to try new things and take risks for the sake of progress. Our government, our legal system, our health system, our educational system: All of these can be radically and greatly improved by the right infostructure, and Americans are more willing and able to push this process ahead than people in other parts of the world.

The critical advances of the next generation involve the development and construction of a radically new infostructure that will change the way government, the law, education, medicine, and many other institutions and industries work. The new infostructures will so dramatically enhance the efficiency of the societies able to build and install them that they will enjoy huge advantages over those that cannot.

America needs to rebuild its infostructure. The alternative to a race to the bottom, to trying to compete on the basis of lower wages or laxer regulations, is a race to the front: to the cutting edge of human progress. This is a race to build the new kinds of infostructure that advances in communications and information technology have made possible. This is a race that we can win and that is worth winning. America can get to a faster, more efficient future faster than other people, and it is on that basis that we can prosper and raise our living standards in the coming years.

Reference

Mead, Walter Russell. 2012. "Beyond the Blue Social Model." *American Interest*, February 4.

Solving the Productivity Puzzle: The Role of Demand and the Promise of Digitization

Jaana Remes

Partner, McKinsey Global Institute

Introduction

Nine years into recovery from the Great Recession, labor productivity-growth rates remain near historic lows in the United States and many other advanced economies. Productivity growth is crucial to increase wages and living standards, and helps raise the purchasing power of consumers to grow demand for goods and services. Therefore, slowing labor productivity growth heightens concerns at a time when aging economies depend on productivity gains to drive economic growth.¹¹⁶ Yet in an era of digitization, with technologies ranging from online marketplaces to machine learning, the disconnect between disappearing productivity growth and rapid technological change could not be more pronounced.

After a year-long research effort at the McKinsey Global Institute, we shed light on the recent slowdown in labor productivity growth in the United States and Western Europe and outline prospects for future growth (see **Box 1**, “Our Methodology”). We find that three waves collided to produce a productivity-weak but job-rich recovery: the waning of a productivity boom that began in the 1990s, financial crisis aftereffects, including weak demand and uncertainty, and digitization. The first two waves have dragged down productivity growth by 1.9 percentage points on average across countries since the mid-2000s, from 2.4% to 0.5%. In particular, financial crisis aftereffects include weak demand, uncertainty, excess capacity, contraction

Reprinted from the McKinsey Global Institute, February 2018, with the permission of the publisher, McKinsey & Company.

This paper is based on research conducted with my coauthors James Manyika, Jacques Bughin, Jonathan Woetzel, Jan Mischke, and Mekala Krishnan, together with our research team, all at the McKinsey Global Institute. For more on the underlying research, see <https://www.mckinsey.com/global-themes/meeting-societys-expectations/solving-the-productivity-puzzle>.

¹¹⁶*Global Growth: Can Productivity Save the Day in an Aging World?* (McKinsey Global Institute, January 2015). Please note that in this report, we often refer to labor productivity as simply “productivity”; we specify other types of productivity, such as total factor productivity, when referring to them.

and expansion of hours, and, in some sectors, a boom-bust cycle. The third wave, digitization, is fundamentally different from the first two because it contains the potential to reignite productivity growth but the benefits have not yet materialized at scale. This is due to adoption barriers and lag effects as well as transition costs. As financial crisis aftereffects recede and more companies incorporate digital solutions, we expect productivity growth to recover; the good news is that we are seeing an uptick today in economic variables like productivity and GDP growth across many countries. We calculate that the productivity-growth potential could be at least 2% per year across countries over the next decade. However, capturing the productivity potential of advanced economies may require a focus on promoting both demand and digital diffusion in addition to more traditional supply-side approaches. Furthermore, continued research will be needed to better understand and measure productivity growth in a digital age.

Box 1. Our Methodology

We analyze the productivity-growth slowdown across a sample of seven countries: France, Germany, Italy, Spain, Sweden, the United Kingdom, and the United States. These countries were chosen to cover a large and diverse portion of GDP in advanced economies, representing about 65%. We do not include any analysis of emerging markets, which have a different productivity-growth dynamic compared to mature markets. In addition to country aggregate analysis, we analyze six sectors across our sample of economies to identify what patterns are similar across sectors and what features are sector-specific, in order to understand what drives aggregate productivity trends. We chose these sectors—automotive manufacturing, finance, retail, technology, tourism, and utilities—because they represent a large and diverse share of the economies in our sample countries and played a significant role in explaining the recent slowdown. In our analysis across countries and sectors, we assess the evidence for today’s leading explanations for the productivity-growth slowdown.¹¹⁷ We find evidence of a non-measurement-related productivity-growth slowdown and therefore

¹¹⁷These include mismeasurement; financial crisis-related factors such as weak investment postcrisis and the rise of zombie firms; and structural shifts such as the rate of technological diffusion, the increasing concentration of businesses, and declining business dynamism together with a growing divergence of productivity among firms, a mix shift toward less productive sectors, a maturation of global supply chains, and secular stagnation.

focus our work in this report on explaining the productivity slowdown as measured.¹¹⁸

We take an integrated analytical approach across supply and demand to assess the linkages and “leakages” around the virtuous cycle of economic growth (from production of goods and services, leading to incomes for households and profits for companies, in turn resulting in continued demand for goods and services). This allows us to diagnose why productivity growth has slowed, particularly as many of the leading explanations today take a supply-focused view rather than an integrated one. In our analysis, we often compare the turn of the century (2000–2004)—a five-year period before the start of the recent productivity-growth slowdown in the United States that encompasses the late boom of 2000, recession of 2001, and recovery period—with the postrecession years (2010–2014), a somewhat stable period a decade later (though encompassing the double-dip recession in Europe). Looking closely at the recent slowdown allows us to identify short-term factors behind the productivity-growth slowdown that are likely to be resolved, as well as long-term trends that are likely to remain in place, helping us to determine the potential for productivity growth in the future.¹¹⁹

While our methodology allows us to provide a much better understanding of the productivity-growth slowdown and the implications for the future, questions for further research surely remain, such as how to better measure the digital economy and understand the economic impact of digital transitions.

Productivity Growth Remains Near Historic Lows, Following a Job-Rich, Productivity-Weak Recovery

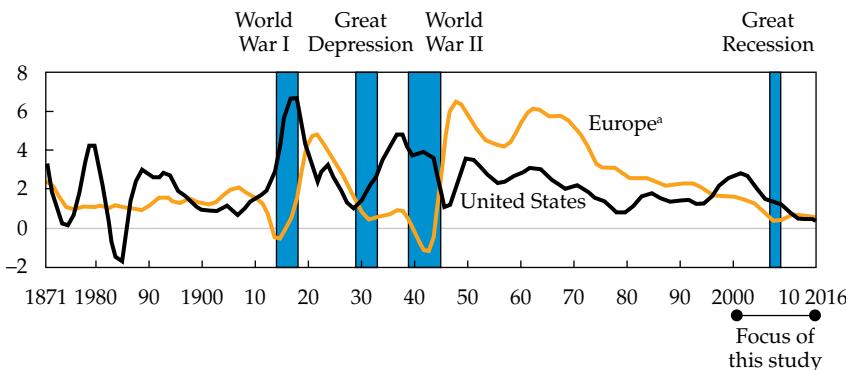
While labor productivity growth has been declining across the United States and Western Europe since a boom in the 1960s, it decelerated further after the financial crisis to historic lows (Figure 1). We focus this study on the slowdown since the early 2000s and identify three major patterns of the productivity-growth slowdown across our sample of countries: low “numerator”

¹¹⁸For more details, see the original publication, Chapter 1, Box 3, “How Significant Could the Mismeasurement of Productivity Growth Be?”

¹¹⁹While we are aware that choosing specific years involves some degree of arbitrariness, after assessing the pros and cons of multiple periods, we determined that concentrating on the period following the crisis allowed us to isolate different factors at the sector level across many different countries more easily. We also conduct robustness tests to assess how much these years impact our results.

Figure 1. Productivity Growth Has Fluctuated over Time; It Has Been Declining since the 1960s and Today Stands Near Historic Lows

Trend Line of Labor Productivity Growth,
Total Economy (% year-over-year)



^aSimple average of France, Germany, Italy, Spain, Sweden, and the United Kingdom.

Note: Productivity defined as GDP per hour worked. Calculated using Hodrick Prescott filter. Drawn from similar analysis in Martin Neil Baily and Nicholas Montalbano, *Why Is Productivity Growth So Slow? Possible Explanations and Policy Responses* (Washington, DC: Brookings Institution, September 2016).

Sources: A. Bergeaud, G. Cette, and R. Lecat, "Productivity Trends in Advanced Countries between 1890 and 2012," *Review of Income and Wealth* 62, no. 3 (2016); McKinsey Global Institute analysis.

(value-added) growth accompanied by robust "denominator" (hours worked) growth, creating a job-rich but productivity-weak recovery across most countries; too few and too small "jumping" sectors; and the critical importance of declining capital intensity growth across countries (see **Box 2**, "Patterns of the Productivity-Growth Slowdown"). These patterns indicate that the productivity-growth slowdown is broad-based across countries and sectors, point to a set of common, overarching factors at work, and reveal the importance of demand-side as well as supply-side factors.

While we find many similar patterns of the productivity-growth slowdown across our sample of countries, there are also notable differences. Sweden and the United States experienced a strong productivity boom in the mid-1990s and early 2000s followed by the largest productivity-growth decline, and much of that decline predated the financial crisis. France and Germany started from more moderate levels and experienced less of a productivity-growth decline, with most of the decline occurring after the crisis. Productivity growth was close to zero in Italy and Spain for some time well before the crisis, so severe labor shedding after the crisis actually accelerated productivity growth.

While many key economic variables such as GDP growth and investment as a share of GDP, as well as productivity growth, have started to pick up recently in the United States and Europe, productivity growth remains low relative to historical levels, with many countries in our sample seeing around 1% productivity growth or less.

Box 2. Patterns of the Productivity-Growth Slowdown

Any explanation of the productivity puzzle should take into account the micro patterns of the slowdown and not just the headline aggregate productivity numbers. We find three major micro patterns. First, the recovery from the financial crisis has been “job-rich” and “productivity-poor” with low “numerator” (value-added) growth accompanied by robust “denominator” (hours worked) growth (Figure 2).¹²⁰ The broad-based pattern of job-rich but productivity-weak recovery across most countries raises the question of why companies are increasing employment without corresponding increases in productivity growth. It also highlights the importance of examining demand-side drivers for slow value-added growth and low productivity growth.

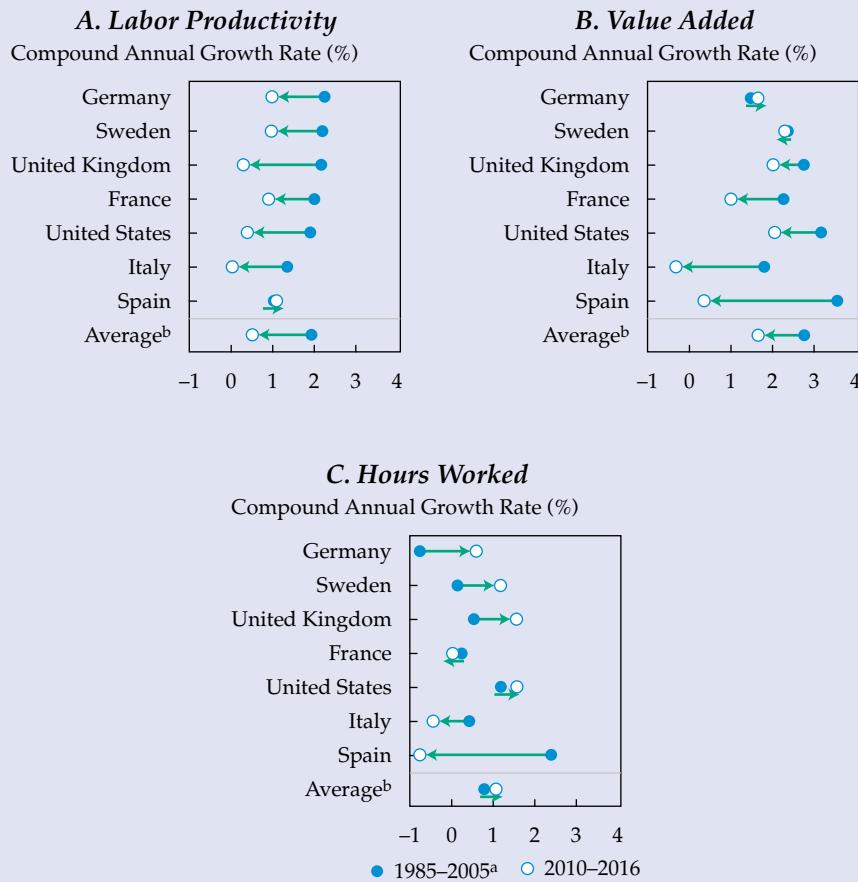
Second, looking across more than two dozen sectors, we find few “jumping” sectors today, and the ones that are accelerating are too small to have an impact on aggregate productivity growth.¹²¹ For example, only 4% of sectors in the United States were classified as jumping in 2014, compared with an average of 18% over the last two decades, and they contributed only 4% to added value.¹²² The distinct lack of jumping sectors we have found across countries is consistent with an environment in which digitization and its benefits to productivity are happening unevenly.

¹²⁰That is not to say economies experienced a jobs boom but that solid job growth continued over a long time through and beyond the period from 2010 to 2014. While some considered this recovery “jobless” early on (see, for example, Natalia A. Kolesnikova and Yang Liu, *Jobless Recoveries: Causes and Consequences*, Federal Reserve Bank of St. Louis, 2011), because it took so long for unemployment to recover, we find that hiring has been exceptionally steady over a long period. The time periods in this exhibit were chosen to allow us to compare a long-term trend (1985 to 2005, ending prior to the crisis, to eliminate the impact of the crisis) with the most recent trends in the recovery (the period of the particularly low productivity growth).

¹²¹A sector is classified as “jumping” in year Y if its compound annual growth rate of productivity for years Y-3 through Y is at least 3 percentage points higher than it was for 1995 to 2014 as a whole (a “long-term” average).

¹²²Similar trends are also seen in Europe. Less than 5% of sectors in France, Germany, Sweden, and the United Kingdom are classified as jumping today.

Figure 2. In Many Countries, Exceptionally Low Productivity-Growth Postrecession Reflects Slowing Value-Added Growth with Robust Growth in Hours Worked



^aLooking at these periods allows us to identify short-term factors behind the productivity-growth slowdown that are likely to be resolved, and long-term trends that are likely to remain in place.

^bWeighted average across France, Germany, Italy, Spain, Sweden, United Kingdom, and United States, based on 2016 GDP (2016 \$ million).

Note: Order of countries is based on fastest to slowest productivity growth in the 1985–2005 period.

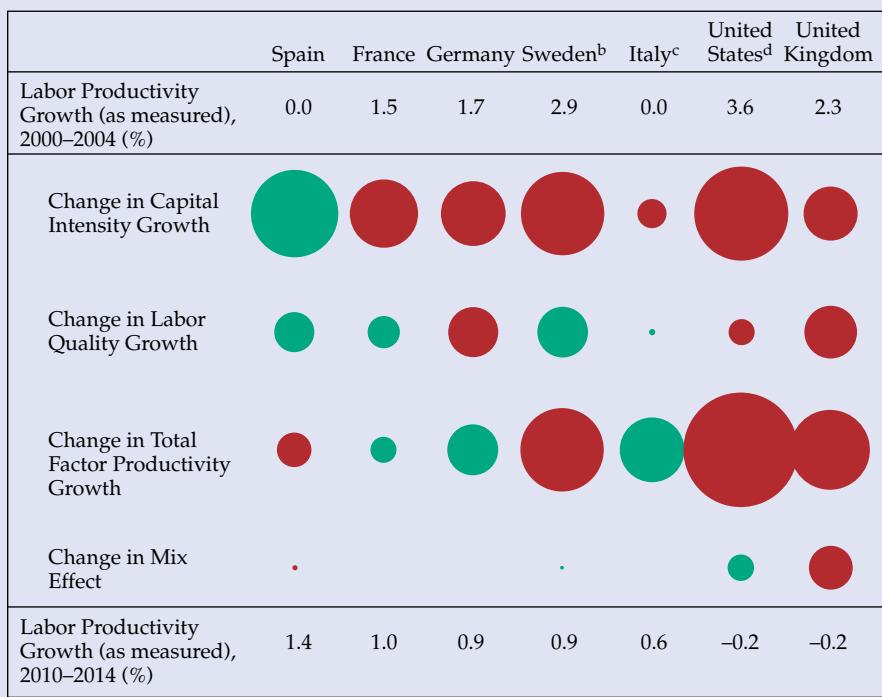
Source: The Conference Board (May 2017 release); McKinsey Global Institute analysis.

Figure 3. Slow Productivity Growth Was Accompanied by a Decline in Capital Intensity Growth, as well as Declining Total Factor Productivity Growth in Some Countries

Contribution to the Decline in Productivity Growth,

2000–2004 vs. 2010–2014^a

Percentage Points



● Increases Productivity Growth ● Decreases Productivity Growth

 Size of Bubble = 0.5

^aSee technical appendix in the original publication for details on methodology.

^bEU KLEMS data on total factor productivity (TFP) was significantly different compared with other data sources such as The Conference Board and Penn World Tables. Hence, we take the average TFP of the three databases and calculate labor quality as a residual.

^cIn Italy, the period analyzed is 2010–2013 instead of 2010–2014 due to data limitations.

^dUS data are for the private business sector only; Europe data are for the total economy.

Note: Order of countries is based on fastest to slowest productivity growth in 2010–2014.

Source: EU KLEMS (2016 release); BLS Multifactor Productivity database (2016 release); McKinsey Global Institute analysis.

Third, since the Great Recession, capital intensity, or capital per hour worked, has in many developed countries grown at the slowest rate in postwar history. Capital intensity indicates access to machinery, tools, and equipment and is measured as capital services per hour. An important way productivity grows is when workers have better tools such as machines for production, computers and mobile phones for analysis and communication, and new software to better design, produce, and ship products, but this has not been occurring at past rates. A decomposition of labor productivity shows that slowing growth of capital per hour worked contributes about half or more of the productivity-growth decline in many countries (**Figure 3**).¹²³

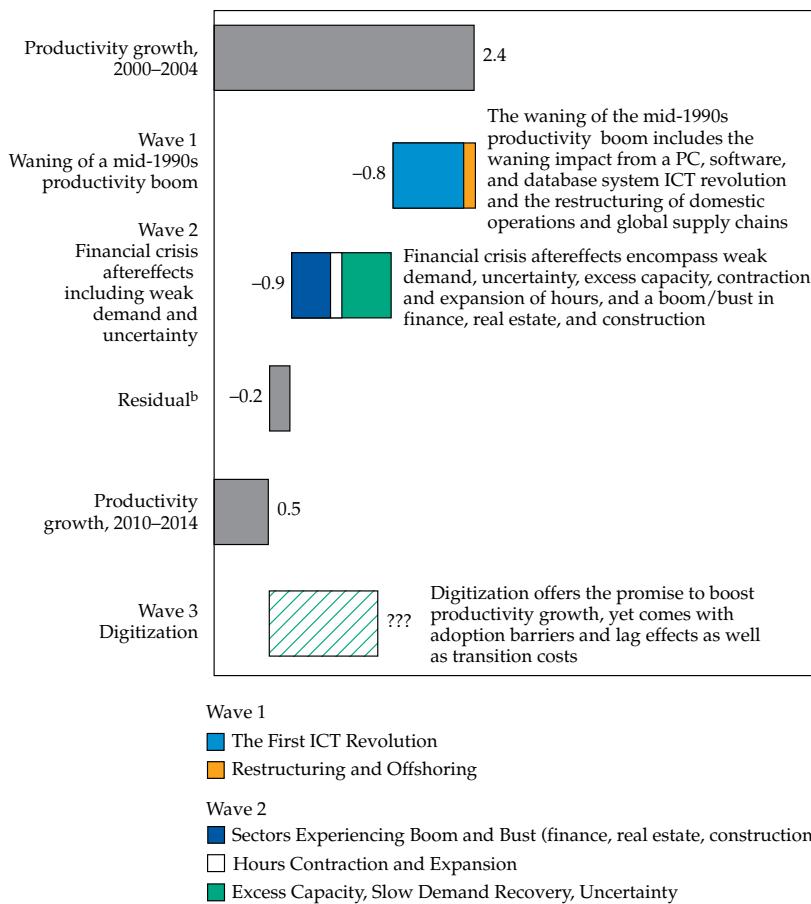
The Waning of a Boom Starting in the 1990s and Financial Crisis Aftereffects Have Dragged Down Productivity Growth to Historic Lows While Digitization Is Under Way

Two waves have dragged down productivity growth by 1.9 percentage points on average across countries since the mid-2000s: the waning of a boom that began in the 1990s with the first information and communications technology (ICT) revolution, and a subsequent phase of restructuring and offshoring, which reduced productivity growth by about 1 percentage point. Financial crisis aftereffects, including weak demand and uncertainty, reduced it by another percentage point. A third wave, digitization, contains the promise of significant productivity-boosting opportunities but the benefits have not yet materialized at scale. This is due to adoption barriers and lag effects as well as transition costs; the net effect on productivity in the short term is unclear (**Figure 4**). We do not attempt to quantify the impact of digitization. Today we find that companies are allocating substantial time and resources to

¹²³We acknowledge that this analysis represents a decomposition and is not a causal analysis, and it is sensitive to the underlying growth accounting formulation. The choice of time periods reflects both the specific trends we want to highlight and constraints from data availability. Comparing the productivity growth in the 2000–04 period with the recent slowdown (2010–14 period) allows us to identify short-term factors behind the productivity-growth slowdown that are likely to be resolved, helping to determine the potential for productivity growth in the future. We were also constrained by a longer-term comparison due to data availability issues across countries in EU KLEMS. Other researchers have also found large contributions from capital intensity growth and total factor productivity growth in the United States; see, for example, Alexander Murray, *What Explains the Post-2004 US Productivity-Growth Slowdown?* (CSLS Research Report 2017-05, 2017).

Figure 4. The Waning of a Mid-1990s Productivity Boom and Financial Crisis Aftereffects Have Contributed Roughly Equally to the Decline in Productivity Growth

Contribution to the Decline in Productivity Growth in 2010–2014 vs. 2000–2004^a
Simple Average of France, Germany, Sweden, United Kingdom, and United States
Percentage Points



^aUS data are for the private business sector only; Europe data are for the total economy.

^bIncludes impact of reallocation (share of total labor and relative price movement) across sectors (“mix effect”) and sectors not considered in our analysis. May include some of the impact from transition costs of digital.

Notes: Italy and Spain are excluded from this analysis because their productivity growth between these time periods did not decrease. We did not attempt to size the impact from Wave 3 (Digitization). While digitization contains the promise of significant productivity-boosting opportunities, it comes with lag effects and adoption barriers as well as transition costs. The net effect on productivity is unclear. Numbers may not sum due to rounding.

Sources: EU KLEMS (2016 release); BLS Multifactor Productivity database (2016 release); McKinsey Global Institute analysis.

changes and innovations that do not yet have a direct and immediate impact on output and productivity growth.

The importance of these waves was not equal across countries. The first wave mattered more in Sweden and the United States, where the productivity boom had been more pronounced, while financial crisis aftereffects were felt more broadly across countries.¹²⁴

Coming into the Crisis, a Boom That Began in the 1990s with the First ICT Revolution and a Restructuring and Offshoring Phase

Waned. An initial ICT-enabled productivity boom, starting in the second half of the 1990s, was particularly strong in Sweden and the United States. The productivity boom in the ICT sector itself reflected a wave of rapid innovation in semiconductor design and manufacturing processes that raised productivity in the sector significantly and translated into higher-quality and higher-value products of downstream computer equipment producers. It also benefited sectors like retail, as large-format retailers like Walmart used technology to transform supply chains and the rest of the industry followed.¹²⁵ The global industry restructuring following the 2001 tech downturn helped sustain productivity gains across manufacturing as production shifted to Asia and nearshore assembly locations in Mexico and Eastern Europe, and manufacturing production employment declined in the United States and Western Europe. In addition, rapid declines in ICT equipment prices encouraged an investment boom in other sectors such as professional and business services, as well as strong growth in the ICT services and software sector, and boosted productivity growth as these industries integrated new technology into their business processes and systems.

By the mid-2000s, the productivity-growth benefits from that first wave of ICT innovation had matured. The retail and wholesale supply chain revolution had largely run its course.

Productivity growth in the tech sector itself declined by roughly 14 percentage points in the United States from 2000–2004 to 2010–2014. The composition of the tech industry had shifted toward skilled labor intensive, less scalable software services. And tech manufacturing became more fragmented and innovation more complex as the proliferation of electronic devices and

¹²⁴This analysis ends at 2014 due to lack of data availability across countries after that date. Please note that this analysis is based on sector-level data. Firm-level trends, some of which we discuss throughout the report, can also play a role in influencing productivity growth.

¹²⁵*How IT Enables Productivity Growth: The US Experience across Three Sectors in the 1990s* (McKinsey Global Institute, November 2002).

applications broadened the demands on performance beyond just processor speed.¹²⁶ For example, the shift in demand toward smartphones requires managing sometimes dozens of sensors from fingerprint recognition and GPS to multiple cameras, all requiring efficient power consumption to save battery time. Virtual world gaming, artificial intelligence, and autonomous driving have dramatically expanded the performance demands on graphics processor units (GPUs). The breadth and depth of innovation is vast, making it harder both to accurately measure improvements and to achieve the past pace of improvements, given that the scale in many specialized chips is lower and cost declines slower.¹²⁷

At the same time, the productivity gains from globalization and offshoring as well as efficiency gains from restructuring post-2001 were plateauing. While we found this trend had a smaller impact on productivity growth across countries than the waning of the ICT-enabled boom, it did affect certain sectors. In the auto sector in the United States, the productivity improvements from restructuring and job declines after the 2001 downturn and of regional footprint optimization across NAFTA tapered off by the mid-2000s. In Germany, regional offshoring to Eastern European countries continues today.

Financial Crisis Aftereffects, including Weak Demand and Heightened Uncertainty, Created a Dynamic of Declining Productivity Growth.

Demand. Demand for goods and services across countries and industries dropped sharply during the financial crisis as people lost jobs, income contracted, and the credit impulse reversed.¹²⁸ For example, in the United States, light-vehicle production fell by 47% between 2007 and 2009 (data from IHS

¹²⁶Some researchers also question whether Moore's law still holds, or takes more effort. See, for example, Kenneth Flamm, "Has Moore's Law Been Repealed? An Economist's Perspective," *Computing in Science and Engineering*, IEEE, 19, no. 2 (2017); Nicholas Bloom et al., "Are Ideas Getting Harder to Find?" (NBER Working Paper 23782, September 2017); and *Moore's Law: Repeal or Renewal?* (McKinsey & Company, 2013).

¹²⁷Other research has also pointed to the importance of the waning of this first ICT-enabled boom. See, for example, John Fernald and Bing Wang, "The Recent Rise and Fall of Rapid Productivity Growth," *FRBSF Economic Letter* (Federal Reserve Bank of San Francisco, September 2015). Others have questioned whether mismeasurement could explain the productivity-growth decline, given the exceptionally thorny challenges of measuring output of rapidly changing tech industries. For a good overview, see David Byrne, Stephen Oliner, and Daniel Sichel, "Prices of High-Tech Products, Mismeasurement, and Pace of Innovation" (NBER Working Paper 23369, April 2017).

¹²⁸The credit impulse is measured as acceleration or deceleration in debt/GDP ratios and thus indicative of the role of borrowing in impacting demand. See, for example, Michael Biggs and Thomas Mayer, *Bring Credit Back into the Monetary Policy Framework!* (Political Economy of Financial Markets policy brief, August 2013).

Markit, 2017 for light-vehicle production), while in retail demand growth slowed by roughly 1 percentage point compared with the pre-crisis period (data from BLS).

This fall in demand for goods and services resulted in significant excess capacity and a pullback of investment. At the same time, in many countries, companies reacted to the demand shock by cutting hours worked, particularly in sectors like manufacturing, retail, finance, and construction. The contraction of hours was so dramatic in the United States that it briefly increased productivity growth in 2009 and 2010.

By the end of 2009, the crisis reached a turning point, with GDP levels bottoming out in the United States. However, the depth of the crisis, deleveraging by households and corporations, weak animal spirits, and structural demand drags such as rising inequality and declining labor share of income resulted in a prolonged recovery that by some measures continues today.¹²⁹ Some European countries also experienced double-dip recessions in 2011.

A combination of factors in this slow recovery period created a dynamic of declining productivity growth: a slow increase in demand, excess capacity, and economic, political, and regulatory uncertainty, all in an environment of low wage growth. This cocktail contributed to the trend of weak growth in productive capital coupled with a rebound in hours worked growth. The decline in the growth rate of capital intensity, the lowest in the postwar period, reflects a substantial decline in equipment and structures investment during the crisis with a slow recovery while intangible investment, such as R&D and software, recovered more quickly after a brief and smaller dip in 2009.¹³⁰ As hours worked had significantly contracted during the crisis and capacity was underutilized, companies met slowly rising demand by filling excess capacity and adding hours. For example, in the auto sector in the United States, growth in hours worked surged after 2010, but total hours still remained below pre-crisis levels in 2015 (based on data from the Bureau of Labor Statistics). Capacity utilization in the United States in light-vehicle production began a significant drop in 2007 and fell by 32 percentage points between 2007 and 2009.¹³¹ From 2009–2010, US capacity

¹²⁹See the original publication, Chapter 1, Box 2, “How the Great Recession Was Different.”

¹³⁰Companies typically see R&D investment as longer term. In many industries with rapid speed of technological change, competitive pressure kept investment a priority for companies.

¹³¹Capacity utilization based on straight-time capacity at a one-, two-, or three-shift/three-crew structure dependent on the shift structure in a plant in a given year. This is calculated without overtime. Based on data from IHS Markit, 2017, for light-vehicle production capacity.

utilization rates went up 20 percentage points, then 9 percentage points from 2010–2011 (thus returning close to 2007 levels), and another 10 percentage points from 2011–2012. Across the economy, once capacity utilization picked up, though, a reason for continued weak investment was the persistent slow recovery in demand. Normalizing capital/output ratios across countries indicates that the investment recovery kept pace with the recovery in demand, but since demand growth was weak, capital services growth remained weak, too.¹³²

Slow wage growth dampened the need to substitute capital for labor. Low wage rates in retail in the United States, for example, seem consistent with comparatively slow investment in technologies like automated check-outs and redeploying freed-up resources in low-productivity occupations like greeters.¹³³ In addition, stagnant wages had implications for limiting demand growth. In our sector analysis, we found weak demand dampened productivity growth through other channels than investment, such as economies of scale and a subsector mix shift (see **Box 3**, “Additional Ways Weak Demand Hurt Productivity Growth during the Recovery”).

¹³²For other explanations of the slowdown, see “Uneven Growth: Short- and Long-Term Factors,” *IMF World Economic Outlook*, (IMF, April 2015); Gustavo Adler et al., “Gone with the Headwinds: Global Productivity” (IMF Staff Discussion Note 17/04, April 2017); “Business Investment Developments in the Euro Area since the Crisis” (ECB Occasional Paper Series, no. 167, European Central Bank, January 2016); Georg Erber, Ulrich Fritzsche, and Patrick Christian Harms, “The Global Productivity Slowdown: Diagnosis, Causes and Remedies,” *Intereconomics* 52, no. 1 (January/February 2017); and Diego Anzoategui et al., “Endogenous Technology Adoption and R&D as Sources of Business Cycle Persistence” (NBER Working Paper 22005, February 2016); Claudio Borio, “Secular Stagnation or Financial Cycle Drag?” (paper presented at National Association for Business Economics 33rd Economic Policy Conference, Washington, DC, 5–7 March 2017); Bart van Ark and Kirsten Jäger, “Recent Trends in Europe’s Output and Productivity Growth Performance at the Sector Level, 2002–2015” (working paper).

¹³³See, for example, Sebastian Vanderzeil, Emma Currier, and Michael Shavel, *Retail Automation: Stranded Workers? Opportunities and Risks for Labor and Automation* (Investor Responsibility Research Center, May 2017). For a review of findings related to the role of minimum wages in impacting employment, see David Neumark, “Employment Effect of Minimum Wages” (IZA Discussion Paper 9715, IZA World of Labor, May 2014). Interestingly, even when retailers are investing in automation, they have tended to move existing workers to other jobs such as food service to keep store service levels up and improve customer engagement.

Box 3. Additional Ways Weak Demand Hurt Productivity Growth during the Recovery

We identify two channels in which weak demand hurt sector productivity growth during the recovery in addition to holding back investment:

Economies of scale. In finance, productivity growth declined, particularly in Spain, the United Kingdom, and the United States, due to contractions in lending volumes that banks were unable to fully offset with staff cuts due to the need for fixed labor (for example, to support branch networks and IT infrastructure). The utilities sector, which has seen flattening demand growth due to energy efficiency policies, as well as a decline in economic activity during the crisis, was similarly not able to downsize labor due to the need to support electricity distribution and the grid infrastructure.

The shape of demand and subsector mix shift. Consumer preferences boosted productivity growth in both the auto and retail sectors from the mid-1990s to the mid-2000s through a shift to higher value-per-unit, more productive goods. Today that trend has slowed. The German and US auto sectors have experienced a trend of customers purchasing higher-value-added SUVs and premium vehicles. This boosted productivity growth by 0.4 to 0.5 percentage point in the auto sector in the early 2000s. That trend has slowed slightly in both countries, contributing only 0.3 percentage point to productivity growth in 2010–2014. Similarly, in retail, we estimate that consumers shifting to higher-value goods, for example, higher-value wines or premium yogurts, contributed 45% to the 1995–2000 retail productivity-growth increase in the United States. This subsequently waned, dragging down productivity growth.

The slow recovery, together with political and regulatory uncertainty in the aftermath of the crisis, may be continuing to restrain investment today.¹³⁴ There is debate around how far the recovery has progressed. For instance,

¹³⁴For a measure of uncertainty, see the Economic Policy Uncertainty Index and see Scott R. Baker, Nicholas Bloom, and Steven J. Davis, “Measuring Economic Policy Uncertainty,” *Quarterly Journal of Economics* 131, no. 4 (November 2016). Research has also shown that long-run uncertainty, which is influenced by policy uncertainty, influences both investment and hiring, but the former is more impacted by it than the latter. This is due to lower depreciation rates and higher adjustment costs of investment relative to hiring. See Jose Maria Barrero, Nicholas Bloom, and Ian Wright, “Short and Long Run Uncertainty” (NBER Working Paper 23676, August 2017).

while we have witnessed an extended period of job growth, employment rates are still well below pre-crisis levels in some countries, notably the United States, where the unemployment rate is around historic lows but labor force participation has not fully recovered.¹³⁵ Household investment remains subdued, and business investment as a share of GDP has only slowly recovered to rates seen before the crisis and has still not fully recovered in parts of Europe. Real investment in structures and equipment remains below trend lines in many countries. Indeed, the latest economic data highlight the fact that capital intensity growth remains noticeably weak across countries. Demand and uncertainty are key drivers. We have found from our global surveys of business that 47% of companies that are increasing their investment budgets are doing so because of an increase in demand, yet 38% of respondents say risk aversion is the key reason for not investing in all attractive opportunities.¹³⁶ However, the good news is that the latest data from Europe and the United States indicate that economic growth is picking up and performance was marginally stronger in 2015–2017 compared with the previous period.

The Benefits of Digitization Have Not Yet Materialized at Scale and Come with Adoption Barriers, Lags, and Transition Costs. While the first wave of ICT investment starting in the mid-1990s was mostly from using technology to deliver supply chain, back-office, and later front-office efficiencies, today we are experiencing a new way of digitization that comes with a more fundamental transformation of entire business models and end-to-end operations. We may be experiencing a renewal of the Solow Paradox of the 1980s, with the digital age around us but not yet visible in the productivity statistics.

There are several reasons that the impact of digital is not yet evident in the productivity numbers. These include lag effects from technological and business readiness to reaching adoption at scale, costs associated with the absorption of management's time and focus on digital transformation, and transition costs and revenue losses for incumbents that can drag sector productivity during the transition; the net impact today of digitization is unclear.¹³⁷ On the

¹³⁵Participation is also low due to long-run trends such as aging; see Danny Yagan, “Employment Hysteresis from the Great Recession” (NBER Working Paper 23844, September 2017).

¹³⁶*McKinsey Quarterly* survey, March 2017.

¹³⁷See also Jacques Bughin and Nicolas van Zeebroeck, “Getting Digital ‘Bucks’: How the Interplay of Disruption and Types of Strategic Responses Shapes Digital Investment Payoffs and Solves the Solow Paradox” (working paper, December 2017); Erik Brynjolfsson, Daniel Rock, and Chad Syverson, “Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics” (NBER Working Paper 24001, November 2017).

lag effects, we have found that digitization has not yet reached scale, with a majority of the economy still not digitized. MGI has calculated that Europe overall operates at only 12% of digital potential, and the United States at 18%, with large sectors lagging in both.¹³⁸ While the ICT, media, financial services, and professional services sectors are rapidly digitizing, other sectors such as education, health care, and construction are not. We also see the lack of scale in our sector deep dives. In retail, for example, we found that the growing share of sales taking place online in the United States added roughly 0.5 percentage point to productivity growth in the sector per year, as those forms of retail are more productive than traditional forms yet those sales are about 10% of retail volume.¹³⁹

History shows that technological diffusion takes time and comes with barriers to adoption.¹⁴⁰ An MGI review of the historical rate of adoption of 25 previous technologies over the past half-century shows that the time from commercial availability to 90% adoption ranges from approximately 8 to 28 years.¹⁴¹ This was demonstrated by the first Solow Paradox of the mid-1970s and 1980s, for example, and the ICT boom in the 1990s. Productivity growth in the United States slowed in the former period, despite innovations

¹³⁸Potential is defined by comparing each sector against a frontier sector defined as the US ICT sector. This analysis uses a set of 18 metrics of digitization spanning assets, usage, and labor. Our use of the term *digitization* and our measurement of it encompasses the digitization of assets, including infrastructure, connected machines, data, and data platforms; the digitization of operations, including processes, payment and business models, and customer and supply chain interactions; and the digitization of the workforce, including worker use of digital tools, digitally skilled workers, and new digital jobs and roles. *Digital Europe: Pushing the Frontier, Capturing the Benefits* (McKinsey Global Institute, June 2016); *Digital America: A Tale of the Haves and Have-Mores* (McKinsey Global Institute, December 2015).

¹³⁹Impact on retail productivity growth calculated based on the mix shift between online and offline retail, assuming today's level of relative productivity between the two segments. Based on data from Euromonitor International, Retailing data (2018 edition) and S&P Capital IQ.

¹⁴⁰Boyan Jovanovic and Peter L. Rousseau, "General Purpose Technologies," in *Handbook of Economic Growth*, vol. 1B, edited by Philippe Aghion and Steven Durlauf, 1181–1224 (New York: Elsevier, 2005). Take the advent of steam power, for example. Productivity growth was quite rapid, at 2% to 3%, when steam power was introduced around 1870 but fell with the arrival of electrification in the 1890s to 1% to 2% in the United States. It was only in the period after 1915, which saw the diffusion of machines operated by stand-alone secondary motors and the widespread establishment of centralized power grids, that electricity finally pervaded businesses and households, and productivity growth began to rise. Then productivity growth rose to 3%. See also Paul David, *Computer and Dynamo: The Modern Productivity Paradox in a Not-Too Distant Mirror*, Warwick Economics Research Paper Series (Coventry, UK: University of Warwick, 1989).

¹⁴¹*A Future That Works: Automation, Employment, and Productivity* (McKinsey Global Institute, January 2017). See also Diego Comin and Bart Hobijn, "An Exploration of Technology Diffusion," *American Economic Review* 100, no. 5 (December 2010): 2031–59.

at the time in the area of microelectronics and communications technology.¹⁴² Productivity gains were not automatic and did not occur in all industries that invested heavily in ICT. Instead, real productivity gains required significant changes in business process, as well as managerial and technical innovation.¹⁴³

The challenge of adoption in the current digital wave may be even harder because of the broad range of uses of digital that not only help improve current processes but fundamentally transform business models and operations. For example, in retail, the first ICT revolution was focused on getting the right goods to the right place at the right time. With digitization, the transition to online requires building a new channel with a new supply chain structure to deliver goods directly to customers and determining what combination of stores and online presence is optimal. Digital also requires significant up-front investment and new skills in data analysis; our survey shows fear of technological obsolescence as well as gaps in digital technical and organizational capabilities as barriers. The current wave of digitization also requires customers to embrace developments such as mobile banking, online shopping, autonomous driving, and resolving questions with a bot. Finally, some incumbents have reasons to actively delay adoption, whether for fear of cannibalization or, in some cases, the challenges of large-scale transformations.

While new digital entrants as well as fast-moving incumbents may increase profits and productivity, others can experience a transition that drags down productivity. As they lose revenue to attackers and their growing digital arms cannibalize revenues further, some companies may end up with duplicate structures and processes, and underutilized capacity in their traditional operations. For example, in retail, when firms increase their online presence and stores or entire malls suffer declining footfall, that cannot readily be remedied.¹⁴⁴ In a recent survey we conducted, companies with digital transformations under way said that 17% of their market share from core products or services was cannibalized by their own digital products or

¹⁴²Paul David, *Computer and Dynamo: The Modern Productivity Paradox in a Not-Too Distant Mirror*, Warwick Economics Research Paper Series (Coventry, UK: University of Warwick, 1989).

¹⁴³*How IT Enables Productivity Growth: The US Experience across Three Sectors in the 1990s* (McKinsey Global Institute, November 2002).

¹⁴⁴See also Jacques Bughin and Nicolas van Zeebroeck, “Getting Digital ‘Bucks’: How the Interplay of Disruption and Types of Strategic Responses Shapes Digital Investment Payoffs and Solves the Solow Paradox” (working paper, December 2017); Jacques Bughin, Laura LaBerge, and Anette Mellbye, “The Case for Digital Reinvention,” *McKinsey Quarterly*, February 2017; and Jacques Bughin and Nicolas van Zeebroeck, “The Right Response to Digital Disruption,” *MIT Sloan Management Review*, April 2017.

services.¹⁴⁵ Industry productivity benefits will then materialize mostly as incumbent businesses restructure or exit, and adoption costs are outweighed by benefits as digitization reaches scale.

Country Variations Provide Greater Insight into the Productivity-Growth Slowdown

We find three broadly similar groups of countries: Sweden, the United Kingdom, and the United States, which have experienced the largest productivity-growth decline in our sample; France and Germany, which experienced a less dramatic drop in productivity growth but a continuing long-term decline; and Italy and Spain, with no decline (Figure 5). These variations are mainly associated with the strength of the boom prior to the financial crisis, the extent of the crisis itself, and differences in labor market flexibility.

A Sector View Highlights a 2% Plus Potential for the Future

Our sector analysis provides an alternative lens to examine the macro trend of declining productivity growth. We find the three waves played out in different ways and to different degrees across sectors. Few sectors illustrate how this perfect storm impacted productivity growth across countries as well as the retail sector. By the time the crisis hit in 2007, the retail sector was at the tail end of a productivity boom that began around 1995. Then weak demand resulting from the financial crisis and recovery made matters worse in two ways: through an overall reduction in sales without a corresponding reduction in labor, and a switch to lower value-per-unit products and brands. As demand began to recover and wages across countries remained low, retailers hired more than they invested. In the middle of this slow recovery and challenging demand environment, the rise of Amazon and the wave of digital disruption occurring in the retail industry added about 0.5 percentage point per year to productivity growth from the shift to more productive online channels, accompanied by transition costs, duplicate structures, and drags on footfall in traditional stores.¹⁴⁶ The tourism sector provides a counterexample. It shows how productivity growth has been slow but steady across many countries from the incorporation of new technology, new business models, increasing consolidation, new competitors, and growing demand.

¹⁴⁵“How Digital Reinventors Are Pulling Away from the Pack,” *McKinsey Digital Global Survey 2017* (McKinsey & Company, October 2017).

¹⁴⁶Impact on US retail sector. Impact calculated based on the mix shift between online and offline retail, assuming today’s level of relative productivity between the two segments. Based on data from Euromonitor International, Retailing data (2018 edition) and S&P Capital IQ.

Figure 5. The Patterns and Factors behind the Productivity-Growth Decline Reveal Similarities and Differences across Countries (percentage points)

		United States ¹	United Kingdom	Sweden	Germany	France	Italy	Spain
Decline in productivity growth ²		-3.8¹	-2.5	-2.0	-0.7	-0.5	0.6	1.4
Low “numerator” (value-added) growth ³		-1.2	-0.8	-0.1	0.1	-1.3	-2.2	-3.2
High “denominator” (hours-worked) growth ³		0.4	1.1	1.1	1.4	-0.2	-0.9	-3.2
Few jumping sectors ⁴		-46¹	-30	-28	-7	-14	0	10
Broad-based productivity-growth decline across sectors ⁵		88¹	87	83	67	70	34	50
Contribution of factors in growth accounting	Capital intensity	-1.5¹	-0.5	-1.2	-0.7	-0.9	-0.2	1.4
	Labor quality	-0.2¹	-0.5	0.5	-0.4	0.2	0.0	0.3
	Total factor productivity	-2.3¹	-1.2	-1.2	0.5	0.2	0.8	-0.2
	Sector mix shift	0.21	-0.4	0.0	-0.1	0.1	0.0	0.0
Impact of waves on productivity growth ²	Waning of a mid-1990s productivity boom	-2.0¹	-0.4	-1.1	-0.2	-0.1	n/a	n/a
	Financial crisis aftereffects	-1.1¹	-1.3	-0.9	-1.2	-0.3	n/a	n/a
Top sectors contributing to the decline in productivity growth	Arts, entertainment, and other services						<input type="checkbox"/>	
	Construction					<input type="checkbox"/>		
	Finance and insurance			<input type="checkbox"/>				
	Information/communication services	<input type="checkbox"/>	<input type="checkbox"/>				<input type="checkbox"/>	
	Manufacturing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Real estate				<input type="checkbox"/>			
	Retail and wholesale	<input type="checkbox"/>		<input type="checkbox"/>				
	Transportation and storage				<input type="checkbox"/>			

Keys: Low or no effect = no treatment. Moderate effect = bold. Large effect = bold and gray.

¹US data are for the private business sector only; Europe data are for the total economy.

²2010–2014 versus 2000–2004.

³2010–2016 versus long-term (1985–2005).

⁴Share of jumping sectors, 2014 versus 2004.

⁵Share of sectors with lower productivity growth in 2010–2014 versus 2000–2004.

Note: Numbers may not sum due to rounding.

Source: McKinsey Global Institute analysis.

As financial crisis aftereffects continue to dissipate, we expect productivity growth to recover from current lows across sectors and countries. Our sector deep dives reveal significant potential to boost productivity growth both from a continuation of more typical productivity opportunities such as operational efficiency gains and from new avenues enabled by digital technologies. Digital automation is just one channel in which digitization will impact productivity growth; digital flows and platforms can also accelerate globalization and global competition, and digital features can substantially increase customer value.¹⁴⁷ Overall, we estimate that the productivity-boosting opportunities could be at least 2% on average per year over the next 10 years, with 60% coming from digital opportunities.¹⁴⁸ While low productivity growth of today may lead to concern about the future, research indicates that past productivity performance is a poor indicator of future productivity growth.¹⁴⁹

The Amplification of Demand Drags and the Potential Industry-Breaking Effects of Digital May Limit the Productivity-Growth

Potential of Advanced Economies. While we found that weak demand hurt productivity growth in the aftermath of the financial crisis, looking ahead, there is concern that some demand drags may be more structural than purely crisis-related. There are several “leakages” along the virtuous cycle of growth (**Figure 6**). Broad-based income growth has diverged from productivity growth, because declining labor share of income and rising inequality are eroding median wage growth, and the rapidly rising costs of housing and education exert a dampening effect on consumer purchasing power.¹⁵⁰ It appears increasingly difficult to make up for weak consumer spending via higher investment, as that very investment is influenced first and foremost

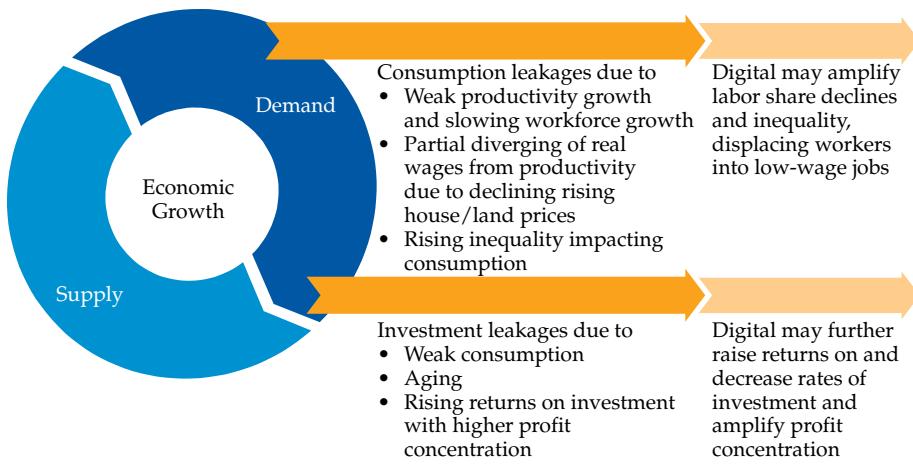
¹⁴⁷*Global Flows in a Digital Age: How Trade, Finance, People, and Data Connect to the World Economy* (McKinsey Global Institute, April 2014).

¹⁴⁸Our estimate for the productivity-growth potential builds on extensive past MGI research on sector opportunities for improving productivity through technologies that are already implemented today or have a clear path to deployment at scale by 2025. These include benefits from digitization (e.g., big data, Internet of Things, automation, AI) as well as non-digital opportunities such as mix shifts in products and channels, continued consolidation, etc. See *Global Growth: Can Productivity Save the Day in an Aging World?* (McKinsey Global Institute, January 2015); *Digital America: A Tale of the Haves and Have-Mores* (McKinsey Global Institute, December 2015); *Digital Europe: Pushing the Frontier, Capturing the Benefits* (McKinsey Global Institute, June 2016); and *A Future That Works: Automation, Employment, and Productivity* (McKinsey Global Institute, January 2017).

¹⁴⁹See Erik Brynjolfsson, Daniel Rock, and Chad Syverson, “Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics” (NBER Working Paper 24001, November 2017).

¹⁵⁰See in the original publication Chapter 5 for a more detailed discussion of declining labor share of income.

Figure 6. Long-Term Demand Leakages Could Act as a Drag on Productivity Growth and May Be Further Amplified by Digital



by demand, and rising returns on investment discourage investment relative to earnings. Demographic trends may further diminish investment needs through an aging population that has less need for residential and infrastructure investment. These demand drags are occurring while interest rates are hovering near the zero lower bound. All of this may hold back the pace at which capital per worker increases, impact company incentives to innovate, and thus impact productivity growth, slowing down the virtuous cycle of growth.

Digitization may further amplify those leakages, for example, as automation may compress labor share of income and increase income inequality by hollowing out middle-class jobs and may polarize the labor market into “superstars” versus the rest. It may also raise returns on investment and thus reduce rates of investment. Cannibalization of incumbent revenues puts pressure on nominal demand. And the rate of technological labor displacement is set to rise. Unless displaced labor can find new highly productive and high-wage occupations, workers may end up in low-wage jobs that create a drag on productivity growth.¹⁵¹ Our ability to create new jobs and skilled workers will impact prospects for income, demand, and productivity growth.

¹⁵¹See also Michael A. Landesmann, and Robert Stehrer, “Technology Diffusion, International Competition and Effective Demand,” *Revue d’économie industrielle* 105, no. 1 (2004); *Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation* (McKinsey Global Institute, November 2017).

Digital technologies may also dampen their own productivity promise through other channels. Various digital technologies are characterized by large network effects, large fixed costs, and close to zero marginal costs. This leads to a winner-take-most dynamic in industries reliant on such technologies and may result in a rise in market power that can skew supply chains and lower incentives to raise productivity. For example, some digital platforms benefit from a growing user base, as social networks with more users allow for more connections, while larger pools of search data generate better and more targeted results. While the potential economic costs and approaches to regulation of network industries are well established, the nature of digital platforms is sufficiently different to warrant further policy consideration.

Independent of platform economies, rising corporate concentration throughout the economy may reduce competitive pressure and translate into weaker incentives to innovate and invest in raising productivity, although we have not found evidence of that yet. While the empirical evidence suggests that the link between concentration and either competitive intensity or productivity growth may not be a strong one, this is another often-cited concern today.¹⁵² Importantly, in our sector deep dives, we have found no evidence that rising business concentration has hurt productivity growth so far. However, going forward, that may not be the case. There may be a tipping point where the initial benefits from industry consolidation, from factors such as economies of scale and reducing the need for staff, and from restructuring operations may give way to costs as competitive pressure declines with the rise of market power. Rising corporate concentration could also further increase income inequality and compress labor share of income.

New digitally enabled business models can also have dramatically different cost structures that change the economics of industry supply significantly and raise questions about whether the majority of companies in the industry and the tail will follow the frontier as much as in the past. For example, in retail, productivity growth in the late 1990s and early 2000s was driven by Tier 2 and 3 retailers replicating the best practices of frontier firms like Walmart. Today, it is unclear if many of Amazon's practices can be replicated by most other retailers, given Amazon's large platform and low marginal cost of offering additional products on its platform. On the other hand, platforms like Amazon, TripAdvisor, and Airbnb offer the potential for new, small, and niche players to compete effectively with larger players, fundamentally

¹⁵²Jason Furman and Peter Orszag, "A Firm-Level Perspective on the Role of Rents in the Rise in Inequality" (presentation at Columbia University, October 2015); Germán Gutiérrez and Thomas Philippon, "Declining Competition and Investment in the US" (NBER Working Paper 23583, March 2017).

changing the structure of the industry. It is unclear, then, what the net productivity impact of such changes in industry structure and economics will be, depending, for example, on the share of the market different players are able to gain and their relative productivity levels.

Finally, digitization may reduce price transparency and market efficiency as the customization of price, product, and terms proliferates through the use of consumer data, potentially reducing the incentives for companies to focus on efficiency gains as they extract more of the consumer surplus.

A New Paradigm for Policy in a Digital Age May Be Warranted.

Unlocking the productivity potential of advanced economies may require a focus on promoting both demand and digital diffusion, in addition to interventions that help remove traditional supply-side constraints such as red tape.¹⁵³ To incentivize broad-based change, companies need competitive pressure to perform better, a business environment and institutions that enable change and creative destruction, and access to infrastructure and talent. Yet additional emphasis on digital diffusion and demand is warranted.

There are many opportunities today for policymakers to help boost productivity growth in advanced economies that focus on demand and digital diffusion. Demand may deserve attention to help boost productivity growth not only during the recovery from the financial crisis but also in terms of longer-term structural leakages and their impact on productivity. Suitable tools for this longer-term situation include focusing on productive investment as a fiscal priority, growing the purchasing power of low-income consumers with the highest propensity to consume, unlocking private business and residential investment, and supporting worker training and transition programs to ensure that periods of transition do not disrupt incomes.

On digital, action is needed both to overcome adoption barriers of large incumbent business and to broaden the adoption of digital tools by all companies and citizens. Actions that can promote digital diffusion include leading by example and digitizing the public sector, leveraging public procurement and investment in R&D, driving digital adoption by small and medium-sized enterprises (SMEs), investing in hard and soft digital infrastructure and clusters, doubling down on the education of digital specialists as well as consumers, ensuring global connectivity, and addressing privacy and cybersecurity

¹⁵³William Lewis, *The Power of Productivity: Wealth, Poverty, and the Threat to Global Stability* (Chicago: University of Chicago Press, 2004); John B. Taylor, “Slow Economic Growth as a Phase in a Policy Performance Cycle,” *Journal of Policy Modeling* 38, no. 4 (July–August 2016); Steven J. Davis, “Regulatory Complexity and Policy Uncertainty: Headwinds of Our Own Making” (paper prepared for Hoover Institution conference on “Restoring Prosperity,” Stanford University, Stanford, CA, February 9–10, 2017).

issues. Furthermore, regulators and policymakers will need to understand the differences in the nature of digital platforms and networks from the network industries of the past and develop the tools to identify non-competitive behavior that could harm consumers.

Other stakeholders have a role to play, too. How do companies, labor organizations, and even economists respond to the challenge of restarting productivity growth in a digital age? Companies will need to develop a productivity strategy that includes the digital transformation of their business model as well as their entire sector and value chain, and not just focus on operational efficiency. In addition, they may have to rethink their employee contract in order to develop a strategy, potentially together with labor organizations, where people and machines can work side by side and workers and companies can prosper together.

Economists can play a key part by developing new and improved ways to measure productivity and by developing models that can assess the impact of technology on markets and prices.

Conclusion

While productivity growth in advanced economies has been slowing for decades, the sharp downturn following the financial crisis has raised alarms. We find that the most recent slowdown is the product of two waves, the waning of a 1990s productivity boom and financial crisis aftereffects, while a third wave, digitization, is under way. As financial crisis aftereffects continue to recede and digitization matures, productivity growth should recover from historic lows. How strong the recovery is, however, will depend on the ability of companies and policymakers to unlock the benefits of digitization and promote sustained demand growth. There is a lot at stake. A dual focus on demand and digitization could unleash a powerful new trend of rising productivity growth that drives prosperity across advanced economies for years to come.

Building a Robust Apprenticeship System in the United States: Why and How?

Robert I. Lerman

Professor of Economics, American University, and a Senior Fellow, the Urban Institute

Skills drive productivity, competitiveness, and incomes. Economic growth is heavily dependent on the growth in human capital (Hanushek and Woessmann 2015). But what is human capital, and what for that matter do we mean by “skills”? Too often, US researchers have identified skills with two key measures: (1) academic attainment in terms of completion of schools and degrees and (2) test scores on academic tests, usually tests of math and verbal capability. Any shortfall in these measures suggests the need for remedial action to help young people extend and complete schooling and to teach them better math and verbal capacities (Goldin and Katz 2008). This consensus view on skills is one reason spending on postsecondary education has grown rapidly and reached record levels per student. The National Center for Education Statistics reports that at the postsecondary level, the United States spent \$27,900 per full-time equivalent student, 89% higher than the OECD average of \$14,800. Now, after mountains of student debt and enormous spending by federal, state, and local governments, the United States is said to face a serious skills mismatch in various occupations, especially those in technical fields. Some academics, consulting firms, and managers see weak skills of many American workers leading to skill shortages and limited economic growth (Deloitte 2011; Carnevale, Smith, and Strohl 2010). One striking indication of a skills gap or mismatch is that German companies operating in the United States identify job skills as a key challenge to their success in the United States and encouraged the German embassy to start a “Skills Initiative” to identify and share information about best practices in sustainable workforce development. Others assert that skills in the United States are not in short supply (Cappelli 2015; Osterman and Weaver 2014). Unfortunately, debates on the adequacy of skills rarely incorporate an appropriately broad definition of skills. The virtual sole emphasis on academic skills as measured by math and verbal test scores and educational attainment is natural because that’s where the data are. This emphasis fails to recognize that productivity

Prepared for the CFA Society New York Conference “Restoring American Economic Dynamism” on 28 November 2017.

depends at least as much on occupational competencies and employability skills, such as communication, teamwork, allocating resources, problem solving, reliability, and responsibility. The myriad nature of skills raises questions about whether added schooling and a targeted focus on academic test scores are the best ways of upgrading skills. So, too, does the recognition that many young people become disengaged from formal schooling, as reflected in weak high school outcomes and high dropout rates from community colleges. Increasingly, policymakers and policy researchers are recognizing the need to shift from the “academic only” approach to teaching skills in schools. Instead, they see enormous potential in expanding apprenticeship, a model that combines work-based learning, production under a mentor/supervisor, wages, along with related courses. A wide body of evidence suggests that apprenticeships are far more cost effective in teaching skills, especially employability and occupational skills, than pure schooling. In Switzerland, perhaps the leading apprenticeship country, an astounding 95% of 25-year-olds have either a BA level degree or a recognized occupational certification, mainly through apprenticeship. About 70% of Swiss youth take up an apprenticeship, though some go on to university programs later. Apprenticeship systems are one of the few mechanisms for improving both the supply and demand sides of the labor market. They are especially effective in teaching occupational and employability skills. Since classroom learning is applied quickly in real-world settings, workers are more likely to retain academic as well as occupational skills. Employers are more likely to create demanding, high-productivity, and well-paying jobs when they can rely on those completing an apprenticeship to have mastered an array of relevant skills and to have gained experience in using those skills. Apprenticeship expansion has become a bipartisan goal, endorsed and acted upon by President Trump at the beginning of his term and President Obama toward the end of his two terms. The Obama administration allocated \$175 million to 46 apprenticeship initiatives by nonprofits and community colleges. At a White House ceremony in June 2017, President Trump called for expanding apprenticeship. He endorsed a “moonshot” goal proposed by Salesforce CEO Marc Benioff to create 5 million apprenticeships in five years. Achieving 5 million apprenticeships would require a ten-fold increase from today’s 440,000 apprentices in civilian sectors and 95,000 in the military. Reaching the 5 million targets might sound impractical, but it would in fact require only that the United States attain about the same share of apprentices in its workforce that Australia and England have already achieved. The president’s first steps to achieve this goal were signing an executive order titled “Expanding Apprenticeship in America” and nearly doubling the funding for apprenticeships to \$200 million. Recognizing the need

for fundamental reforms in the nation’s apprenticeship system, the executive order establishes a task force to examine other administrative and legislative reforms, strategies for creating industry-recognized apprenticeships, and the best ways to encourage the private sector to create apprenticeships. The newly created task force was chaired by the secretary of labor and co-chaired by the secretaries of education and commerce.¹⁵⁴ Meanwhile, bipartisan bills in Congress have called for providing tax credits to companies that offer apprenticeships. Several Republican and Democratic governors are taking steps aimed at expanding apprenticeships. For example, former Republican Governor Scott Walker recently doubled the funding for Wisconsin’s successful youth apprenticeship program. And former Governor John Hickenlooper, a Colorado Democrat, is playing a leading role in creating a youth apprenticeship program modeled after the Swiss system.

This paper begins by defining apprenticeships and then discusses why apprenticeship can be an especially cost-effective approach to increasing skills, productivity, and ultimately wages. At this point, with the increasing acceptance of the rationale for expanding apprenticeship, the paper turns to the “how” questions. Is it feasible to scale up the US system to reach numbers comparable to those in Australia and England while maintaining high quality? If so, what steps are required to do so?

Defining Apprenticeship and Explaining Its Advantages

Apprenticeship training is a highly developed system for raising the skills and productivity of workers in a wide range of occupations, with demonstrated success abroad and scattered examples of success domestically. Apprentices are employees who have formal agreements with employers to carry out a recognized program of work-based and classroom learning as well as a wage schedule that includes increases over the apprenticeship period. Apprenticeship prepares workers to master occupational skills and achieve career success. Under apprenticeship programs, individuals undertake productive work for their employer; earn a salary; receive training primarily through supervised, work-based learning; take academic instruction that is related to the apprenticeship occupation; and receive a certificate of completion. The programs generally last from two to four years. Apprenticeship helps workers to master not only relevant occupational skills but also other work-related skills, including communication, problem solving, allocating resources, and dealing with supervisors and a diverse set of coworkers. The

¹⁵⁴The task force on apprenticeship expansion released its report to the President on May 10, 2018. See <https://www.dol.gov/apprenticeship/docs/task-force-apprenticeship-expansion-report.pdf>.

coursework is generally equivalent to at least one year of community college. In Austria, Germany, and Switzerland, extensive apprenticeships offer a way of upgrading the quality of jobs, especially in manufacturing, commercial, and managerial positions.¹⁵⁵ In these countries, apprenticeships begin mostly in the late high school years, absorbing 50% to 70% of young people on their way to valued occupational qualifications (Hoffman 2011). OECD reports (2009, 2010) highlight the role of a robust apprenticeship system in limiting youth unemployment. Apprenticeships within the United States and elsewhere show how construction occupations can reach high wages and high productivity. The question is whether the model can be extended and attract firms to upgrade other occupations. Apprenticeship expansion holds the possibility of substantially improving skills and careers of a broad segment of the US workforce. Completing apprenticeship training yields a recognized and valued credential attesting to mastery of skill required in the relevant occupation. Apprenticeships are distinctive in enhancing both the worker supply side and the employer demand side of the labor market. On the supply side, the financial gains to apprenticeships are strikingly high. Studies in the United States indicate that apprentices do not have to sacrifice earnings during their education and training and that their long-term earnings benefits exceed the gains they would have accumulated after graduating from community college (Hollenbeck 2008). The latest reports from the state of Washington show that the gains in earnings from various education and training programs far surpassed the gains to all other alternatives (Washington State Workforce Training and Education Coordinating Board 2014). A broad study of apprenticeship in 10 US states also documents large and statistically significant earnings gains from participating in apprenticeship (Reed et al. 2012).

These results are consistent with many studies of apprenticeship training in Europe, showing high rates of return to workers. One recent study managed to overcome the obstacle that such studies tend to face where unmeasured attributes explain both who is selected for an apprenticeship and how well apprentices do in the labor market (Fersterer, Pischke, and Winter-Ebmer 2008); the authors did so by examining how an event unrelated to the apprenticeship (the firm staying in or going out of business) caused some apprentices to have full apprenticeships whereas others found their apprenticeships cut short. The estimates indicated that apprenticeship training raises wages by about 4% per year of apprenticeship training. For a three- to four-year apprenticeship, post-apprenticeship wages ended up 12% to 16% higher than they otherwise would

¹⁵⁵For a list of occupations using apprenticeships in several countries, see the occupational standards section of the American Institute for Innovative Apprenticeship website at www.innovativeapprenticeship.org.

be. Because the worker's costs of participating in an apprenticeship are often minimal, the Austrian study indicated high overall benefits relative to modest costs. On the demand side, employers can feel comfortable upgrading their jobs, knowing that their apprenticeship programs will ensure an adequate supply of well-trained workers. Firms reap several advantages from their apprenticeship investments (Lerman 2014). They save significant sums in recruitment and training costs and reduced errors in placing employees, avoiding excessive costs when the demand for skilled workers cannot be quickly filled and knowing that all employees are well versed with company procedures. Because employers achieve positive returns to their investments in apprenticeship, the worker and the government can save significantly relative to conventional education and training. After reviewing several empirical studies, Muhlemann and Wolter (2014) conclude that "in a well-functioning apprenticeship training system, a large share of training firms can recoup their training investments by the end of the training period. As training firms often succeed in retaining the most suitable apprentices, offering apprenticeships is an attractive strategy to recruit their future skilled work force." A recent detailed study conducted by the US Department of Commerce and Case Western University (Helper et al. 2016) found 40% to 50% returns to two expensive apprenticeship programs. One benefit to firms rarely captured in studies is the positive impact of apprenticeships on innovation. Well-trained workers are more likely to understand the complexities of a firm's production processes and therefore identify and implement technological improvements, especially incremental innovations to improve existing products and processes. A study of German establishments documented this connection and found a clear relationship between the extent of in-company training and subsequent innovation (Bauernschuster, Falck, and Heblisch 2009). Noneconomic outcomes are difficult to quantify, but evidence from Europe suggests that vocational education and training in general is linked to higher confidence and self-esteem, improved health, higher citizen participation, and higher job satisfaction (Cedefop 2011). These relationships hold even after controlling for income. An Australian study found that quality apprenticeships improve mental health (Buchanan, et al. 2016). In the United States, evidence from surveys of more than 900 employers indicates that the overwhelming majority believe their programs are valuable and involve net gains (Lerman, Eyster, and Chambers 2009). Nearly all sponsors reported that the apprenticeship program helps them meet their skill demands—87% reported they would strongly recommend registered apprenticeships; an additional 11% recommended apprenticeships with some reservations. Other benefits of apprenticeships include reliably documenting appropriate skills, raising worker productivity, increasing worker morale, and reducing safety problems.

While apprenticeships offer a productivity-enhancing approach to reducing inequality and expanding opportunity, the numbers in the United States have declined in recent years to about one-tenth the levels in Australia, Canada, and Great Britain. Some believe the problems are inadequate information about and familiarity with apprenticeship, an inadequate infrastructure, and expectations that sufficient skills will emerge from community college programs. Others see the main problem as an unwillingness of US companies to invest no matter how favorable government subsidy and marketing policies are. In considering these explanations, we should remember that even in countries with robust apprenticeship systems, only a minority of firms hires apprentices. Because applicants already far exceed the number of apprenticeship slots, the main problem today is to increase the number of apprenticeship openings that employers offer. Counseling young people about potential apprenticeships is a sensible complementary strategy to working with the companies, but encouraging interest in apprenticeship could be counterproductive without a major increase in apprenticeship slots. Apprenticeships are a useful tool for enhancing youth development. Unlike the normal part-time jobs of high school and college students, apprenticeships integrate what young people learn on the job and in the classroom. Young people work with natural adult mentors who offer guidance but allow youth to make their own mistakes (Halpern 2009). Youth see themselves judged by the established standards of a discipline, including deadlines and the genuine constraints and unexpected difficulties that arise in the profession. Mentors and other supervisors not only teach young people occupational and employability skills but also offer encouragement and guidance, provide immediate feedback on performance, and impose discipline. In most apprenticeships, poor grades in related academic courses can force the apprentice to withdraw from the program. Unlike community colleges or high schools, where one counselor must guide hundreds of students, each mentor deals with only a few apprentices.

The high levels of apprenticeship activity in Australia, Great Britain, and Canada demonstrate that even companies in labor markets with few restrictions on hiring, firing, and wages are willing to invest in apprenticeship training. While no rigorous evidence is available about the apprenticeship's costs and benefits to US employers, research in other countries indicates that employers gain financially from their apprenticeship investments (Lerman 2014). Although apprenticeship training can prepare workers for a wide range of occupations, including engineering and architecture,¹⁵⁶ apprenticeships are especially appropriate for skilled positions that do not require a B.A. degree.

¹⁵⁶The United Kingdom features an array of apprenticeships with college degrees in a variety of fields. See <https://www.instituteforapprenticeships.org/apprenticeship-standards/?levelFrom=5>.

Are Apprenticeship Skills Portable?

Concerns about whether the skills learned in apprenticeships bring the portability required to adapt to technical changes have recently surfaced (Hanushek et al. 2017). Using cross-country regressions, the authors find countries that emphasize vocational education improve labor market outcomes in the short-run, but not in the long-run. While impacts are likely to vary by occupation, detailed studies indicate a high degree of skill portability associated with apprenticeship training.

To operationalize the concept of skill specificity, Geel and Backes-Gellner (2009) and Geel, Mure, and Backes-Gellner (2011) borrow an insight from Lazear (2009) that all skills are general in some sense, and occupation-specific skills are composed of various mixes of skills. The authors compile the key skills and their importance for nearly 80 occupations. They then use cluster analysis to estimate how skills are grouped within narrow occupations. This approach recognizes that skills ostensibly developed for one occupation can be useful in other occupations. It identifies occupational clusters that possess similar skill combinations within a given cluster and different skill combinations between clusters. Next, indices for each narrow occupation measure the extent to which the occupation is relatively portable between occupations within the same cluster and/or relatively portable between the initial occupation and all other occupations. The authors use these indices to determine how portability affects mobility, the wage gains and losses in moving between occupations, and the likelihood that employers will invest in training.

The authors test their hypotheses based on empirical analyses of German apprentices. One finding is that while only 42% of apprentices stay in their initial occupation, nearly two-thirds remain with either the occupation they learned as an apprentice or another occupation in the cluster using a similar mix of skills. Second, those trained in occupations with more specific skill sets are most likely to remain in their initial occupation or move to occupations within the same cluster. Third, apprentices increase their wages when moving to another occupation within the same cluster but lose somewhat when moving to another cluster. Fourth, as Geel, Mure, and Backes-Gellner (2011) show, employers are especially likely to invest in apprenticeships with the most specific skill sets. Other strong evidence of the high returns and transferability of German apprenticeship training comes from Clark and Fahr (2001). They examine the returns to apprenticeship for those who remain in the original apprentice occupation as well as losses that do or would occur from transferring to another occupation. The overall rates of return to each year of apprenticeship range from 8% to 12% for training in firms of 50 workers or more and from about 5.5% to 6.5% for firms of 2 to 49 workers. Transferring

to another occupation can offset these gains, but the reduction is zero for those who quit and only 1.7% for those who are displaced from their job and shift to another occupation. As found by Geel and Backes-Gellner (2009), the wage penalty varies with the distance from the original occupation. There is no penalty at all from displacement into a somewhat related occupation. Göggel and Zwick (2012) show the net gains or losses from switching employers and occupations differ by the original training occupation, with apprentices in industrial occupations experiencing wage advantages, whereas those in commerce, trading, and construction see modest losses.

Finally, Clark and Fahr (2001) present workers' own views on the extent to which they use the skills learned in apprenticeship training in their current jobs. Not surprisingly, 85% of workers remaining within their training occupation use many or very many of the skills they learned through apprenticeship. This group constitutes 55% of the sample. But even among the remaining 45%, about two of five workers reported using many or very many of the skills from their apprenticeship and one in five used some of the skills. Overall, only 18% of all former apprentices stated they used few or no skills learned in their apprenticeships. The findings show that the skills taught in German apprenticeship training are often general. Even when bundled for a specific occupation, the skills are portable across a cluster of occupations. Moreover, apprentices are quite likely to remain in occupations that use the skills they learned in their initial occupation. Apprenticeship skills do vary in terms of specificity and portability. But when the skills are less portable, firms are more likely to make the necessary investments and workers are less likely to change occupations significantly. The general component of training is presumably stronger in school-based programs because they are financed by government and/or individuals themselves. Yet, it is far from clear that these programs, especially the purely academic tracks in US secondary schools and US community colleges, offer more mobility. A high percentage of students drop out of academic secondary and community college programs. Also, many of the community college programs are at least as specific as apprenticeship programs. Certificate programs within community colleges are almost entirely devoted to learning a narrow occupational skill, such as courses to become a phlebotomist, child care assistant, or plastics-processing worker. Many US school-based programs take place in for-profit colleges offering narrow programs, such as training to drive trucks, become a medical assistant, and engage in medical insurance billing and coding. Furthermore, skills often erode when they go unused. To the extent students learn general skills but rarely apply them and wind up forgetting them, their training is unlikely to offer upward mobility.

While community college and private for-profit students often take highly specific occupational courses, apprentices all take some general classroom courses. Thus, apprentice electricians learn the principles of science, especially those related to electricity. In most countries, collaboration takes place between public vocational schools and apprenticeship programs. In the United States, apprentices often take their required “related instruction” in classes at community colleges or for-profit colleges (Lerman 2010). From this perspective, apprenticeship programs should be viewed as “dual” programs that combine work- and school-based learning, albeit with an emphasis on work-based learning.

Can the United States Scale Up Apprenticeships?

With the desirability of expanding apprenticeships gaining widespread support, the issue is now becoming one of feasibility. Can the United States scale apprenticeships and thereby widen the routes to rewarding careers and raising the quality and productivity of jobs? If so, how? A common argument was that the United States lacked the cultural legacy of guilds common in parts of Europe, especially in Austria, Germany, and Switzerland. Another was that US employers will never invest in the in-depth training of their workers. Both arguments have been weakened by experience. In the last two decades, Australia and England, two Anglo countries without the Continental European cultural legacy, have more than tripled their apprenticeships almost to the proportions of the labor force found in Austria, Germany, and Switzerland. In the United States, South Carolina managed to increase the number of companies adopting apprenticeship programs from 90 in 2007 to over 800 in the subsequent eight years by using a combination of modest funding and a high-quality marketing and sales initiative.

The biggest reason for lacking a robust apprenticeship system in the United States is the failure to try. Today, even after recent allocations demonstration funding, government spending on apprenticeships is minimal compared with spending by other countries as well as compared with what it costs to pay for less effective career and community college systems that provide education and training for specific occupations. While total government funding for apprenticeship in the United States has been only about \$100 to \$400 per apprentice annually, federal, state, and local government spending annually per participant in two-year public colleges was approximately \$11,400 in 2008 dollars (Cellini 2012). Not only are government outlays sharply higher, but the cost differentials are even greater after accounting for the foregone earnings of college students as they learn far exceed any foregone earnings apprentices experience. Nearly all other countries with

significant apprenticeship programs pay for the off-job courses required in an apprenticeship. The United States rarely does so. Until recently, the federal government has been spending less than \$30 million annually to supervise, market, regulate, and publicize the system. Many states have had only one employee working under the Office of Apprenticeship. Were the United States to spend what Britain spends annually on apprenticeship, adjusting for the differences in the labor force, it would provide at least \$9 billion per year for apprenticeship. Note that the federal Pell Grant program for low- and lower-middle-income college students costs about \$33 billion per year, with a good chunk of the spending going toward career-focused programs in community and career colleges. Thus, at least some of the low apprenticeship penetration is attributable to a lack of public effort in promoting and supporting apprenticeship and to heavy subsidies for alternatives to apprenticeship. Still, other barriers to expansion are significant. One is limited information about apprenticeship. Because few employers offer apprenticeships, most employers are unlikely to hear about apprenticeships from other employers or from workers in other firms. Compounding the problem is both the difficulty of finding information about the content of existing programs and the fact that developing apprenticeships is complicated for most employers, often requiring technical assistance that is minimal in most of the country. Another barrier is employer misperceptions that apprenticeship will bring in unions. There is no evidence that adopting an apprenticeship program will increase the likelihood of unionization, but reports about such close links persist. An additional barrier is the asymmetric treatment of government postsecondary funding, with courses in colleges receiving support but courses related to apprenticeship receiving little financial support. Policies to reduce the government spending differentials between college subsidies and apprenticeship subsidies can help overcome this barrier.

Whether to emphasize apprenticeships beginning in late high school or after high school involves trade-offs. High school programs improve the likelihood of government funding for academic courses related to apprenticeships. Given the consensus that the government should fund students through secondary school, paying for the related instruction of high school apprentices becomes a nondiscretionary part of budgets. When apprentices are beyond high school, government funding for related instruction must come out of discretionary expenses. International experience demonstrates the feasibility of youth apprenticeships; youth can attain serious occupational competencies while completing secondary education. Apprenticeships in the late teenage years improve the nonacademic skills of youth at a critical time. In countries with little or no youth apprenticeship, structured

work experience is less common, limiting the ability of youth to develop critical employability skills such as teamwork, communication, problem solving, and responsibility. Early apprenticeships can help engage youth and build their identity (Halpern 2009). Apprentices work in disciplines that are interesting and new; they develop independence and self-confidence through their ability to perform difficult tasks. Youth try out new identities in an occupational arena and experience learning in the context of production and making things. From an economic perspective, apprenticeships for youth can be less costly for employers. Wages can be lower partly because youth have fewer medium- and high-wage alternatives and partly because youth have fewer family responsibilities, allowing them to sacrifice current for future income more easily. Although Swiss firms invest large amounts of money in their apprenticeship programs, they pay their young apprentices very low wages during the apprenticeship period. Another economic advantage is that starting earlier in one's career allows for a longer period of economic returns to training. For the United States, scaling apprenticeship in the last years of high school is difficult. The aversion to tracking students too early into an occupational sequence is a common objection to youth apprenticeship. Important to note is that high school officials are generally averse to adding youth apprenticeship to their already extensive agenda, including implementing Common Core standards and school and teacher accountability standards as well as dealing with charter schools and vouchers. In the early 1990s, opposition to youth apprenticeship in the United States came from unions and others who worried about eroding the apprenticeship brand with less intensive training programs. While the verdict is still out on whether the United States can achieve scale in apprenticeships, its best chance is to assess where the system needs to go and to take incremental steps to get there.

Ten Elements for a Robust Apprenticeship System

Broad political and industry support are necessary but not sufficient to build and sustain a robust apprenticeship system. In addition, several elements are required for the system to work well. These include:

1. Effective branding and broad marketing
2. Incentives for selling and organizing apprenticeships to private and public employers
3. Programs to develop credible occupational standards with continuing research

4. Endpoint assessments of apprentices and programs
5. Certification body to issue credentials
6. Making apprenticeships easy for employers to create and to track progress
7. Funding for quality instruction in off-job classes
8. Counseling, screening prospective apprentices to ensure they are well-prepared
9. Training the trainers for apprenticeship
10. Research, evaluation, and dissemination

Branding Apprenticeship. Recent successes in Britain and South Carolina have been accompanied by a concerted effort to create apprenticeship as a distinctive brand. South Carolina chose to link apprenticeship with local pride with the brand name of Apprenticeship Carolina. Britain began its growth with the name Modern Apprenticeships but subsequently allowed the apprenticeship label to stand on its own while copyrighting the term. It is now illegal to call an employer training program an apprenticeship unless it is under the official apprenticeship system. At the same time, Britain spent millions of dollars advertising apprenticeships, including advertisements on the London subways.

Selling and Organizing Apprenticeships. Branding and broad marketing will not suffice without a well-developed system for selling and organizing apprenticeships. An employer convinced by an advertisement must have a place to call to learn about and implement an apprenticeship in the organization. Britain's success in expanding apprenticeships offers one example of how to create successful national and decentralized marketing initiatives. Alongside various national efforts, including the National Apprenticeship Service and industry skill sector councils, the British government provides funding to private training organizations and to colleges of further education for the off-job instruction in apprenticeships. These funds have been sufficient to encourage these organizations to sell and organize apprenticeships with employers. In fact, the British approach has buttressed an association of private companies that engage in the kind of retail marketing required to persuade employers to offer apprenticeships. Another step is the British government's initiative to create apprenticeships within the civil service, specifying that apprentices should constitute 2.3% of government employment.

The success of South Carolina in selling and organizing apprenticeships has depended on the skills of small staff built originally by Ann Marie Stieritz,

the director of Apprenticeship Carolina. She hired individuals who understand businesses, who are engaging, who had worked in companies, ideally the business services industry, and who knew how to develop and manage relationships. She did not require knowledge or experience of apprenticeship. For the first two weeks, the staff engaged in a total immersion-learning process about apprenticeship, where they learned about the concept of apprenticeship, became familiar with apprenticeship regulations and forms, and saw apprenticeship programs firsthand. The staff worked closely with Ron Johnson, a career employee and the federal Office of Apprenticeship's representative for South Carolina. The presence of Johnson and his flexibility in pushing for the approval of company programs was important in the initiative's ability to expand within the context of the registered apprenticeship system.

The expansion of apprenticeship has involved reaching out across broad industry sectors, including advanced manufacturing, health care, and information technology. Apprenticeship marketing often takes place in the context of state and local economic development efforts to attract new businesses. The program's work with companies on their training needs is marketed as a reason for a firm to locate in South Carolina. Workforce Innovation and Opportunity Act (WIOA) agencies are also cooperating, sometimes providing on-the-job training subsidies in the context of apprenticeship. The South Carolina Chamber of Commerce publicizes apprenticeship through forums, newsletters, and committee meetings. The value added by Apprenticeship Carolina comes mainly from the program's ability to work with businesses to diagnose their skill demands, including what they see as an ideal set of skills they want workers to master.

Credible Occupational Standards. Nearly all countries with robust apprenticeship systems create occupational frameworks for apprenticeship that all employers training in the relevant occupation mainly follow, with modest additions relating to their own organization. The current US "registered apprenticeship" system is unique in requiring individual companies or other sponsors (such as unions) that wish to register their programs to supply their own skill frameworks and curriculum. In half the states, the approval process is subject to the preferences of state agencies that are often highly restrictive and that require excessive numbers of skilled workers/mentors (people who have completed an apprenticeship in the field or have occupational expertise developed elsewhere) per apprentice. Pennsylvania, for example, mandates a ratio of 4 to 1.

The structure for registered apprenticeships in the United States leads to skill frameworks that are often uneven and highly variable. While joint

employer-union construction apprenticeship programs generally use common frameworks for each occupation, even union programs can vary from state to state.

Employers rarely have the time or common vision across employers to develop frameworks on their own. Moreover, the frameworks should reflect the interests of the apprentices as well as the interests of the employers. This is especially the case if the public sector provides some funding for the programs to take account of the general skills (skills that have value outside the training firm) taught.

Countries vary in their approaches, but all rely on the cooperation of the public and private sectors. The Institute for Apprenticeship in England recently began operating, with the responsibility to oversee skill frameworks initially created by leading employers using the occupation. In Switzerland, the Federal Office for Professional Education and Technology, together with cantons, employers, trade associations, and unions, participates in framing the occupational standards for about 250 occupations (Hoeckel, Field, and Grubb 2009). The canton vocational education programs implement and supervise the vocational schools, career guidance, and inspection of participating companies and industry training centers. Professional organizations develop qualifications and exams and help develop apprenticeship places. Occupational standards in Germany are determined primarily by the “social partners,” including government, employer, and employee representatives (Hoeckel and Schwartz 2009). The chambers of commerce advise participating companies, register apprenticeship contracts, examine the suitability of training firms and trainers, and set up and grade final exams.

The content of skill requirements in apprenticeships includes academic courses and structured work-based training. In each field, the requirements are to complete the coursework in a satisfactory manner and to demonstrate the apprentice’s ability to master a range of tasks. Some systems require apprentices to complete a set of general tasks that apply to a family of occupations (say, metalworking) as well as tasks that apply to a specific occupation (say, tool mechanics or metal construction and shipbuilding). Although the tasks vary widely across occupations, all involve the application of concepts and academic competencies.

Under a contract from the US Department of Labor, the Urban Institute in collaboration with the American Institute for Innovative Apprenticeship has begun publishing competency-based occupational frameworks for apprenticeships in several occupations.¹⁵⁷ This approach could form the foundation

¹⁵⁷See <https://innovativeapprenticeship.org/us-apprenticeships/> for examples.

for what President Trump’s executive order calls “industry-recognized apprenticeships.” The idea of moving away from the registered apprenticeship approach of recognizing and registering occupational apprenticeship programs on a company-by-company basis has been criticized of loosening quality standards. The argument is that limiting the government role in approving programs will lower the quality of apprenticeships. In fact, developing broad-based industry standards is likely to raise quality and to move the US system toward national frameworks that are common in all other countries with robust apprenticeship programs.

Assessments and Certifications. The extent to which systems develop third-party assessments varies across countries. In Germany, each apprentice is subject to an examination by six to nine experts in the occupation, including representatives from chambers of commerce and educators. Several organizations in Britain provide what are called end-point assessments as well as interim assessments. One of the largest is City and Guilds, a private organization that provides curricula as well as assessment services for a large number of apprenticeships. In addition, most countries provide audits of overall programs, including on-the-job learning and the quality of off-job-related instruction.¹⁵⁸

In the United States, federal and state offices lack the staff to audit programs for quality or to provide third-party assessments of apprentices. State and federal apprenticeship agencies do award certifications of completion based on employers’ reports of the apprentices’ progress through their programs. Although completion certificates under the state as well as federal registered apprenticeship system are supposed to be portable throughout the United States, not all states recognize completers from state or federal programs they view as subpar.

Making Apprenticeships Easy for Employers to Create. Marketing to firms through existing federal and state agencies has not worked to scale apprenticeships so far. Although the lack of staff and minimal funding for even the off-job components of apprenticeships play major roles, the system’s complexity can also be a barrier. South Carolina’s sales representatives show that it is possible in some contexts to simplify the process of developing an apprenticeship occupational framework and completing all the paperwork necessary to register the program. The state apprenticeship tax credit of \$1,000 per apprentice per year is also simple to claim. However, the case of South Carolina is an exception. One reason is that the absence of common

¹⁵⁸In England, Ofsted, an agency that reports directly to Parliament, rates the quality of apprenticeship providers.

occupational frameworks that are well recognized as yielding quality outcomes. Another is the federal and state approval process. And a third is the absence of talented human resource consultants who can sell and organize apprenticeships.

Funding for Off-Job Classes Related to the Apprenticeships. One can make a strong theoretical and practical case for the training firm not funding the off-job learning in an apprenticeship. Theoretically, the skills learned in the off-job courses are general in the Becker sense that the worker's added productivity can be applied not only to the current employer but to many other employers. For this reason, the employer cannot recoup the provision of this general training. The worker gains the benefit, but the government shares the worker's gain in the form of higher taxes and reduced transfers. On the practical side, the government already funds a significant share of the costs of courses aimed at teaching occupational skills but does so in a way that is far less cost efficient than apprenticeship.

Judging by the case of England, government financing of delivering courses for apprenticeships could be enough to encourage training organizations to sell employers on apprenticeships. Using a pay-for-performance model, technical education and training organizations would earn revenue only for apprenticeships that each college or organization stimulates.

The government could reap savings from this approach, since every apprenticeship slot stimulated by an already funded college or training organization increases the work-based component of training borne by the employer and reduces the classroom-based component often borne by government. Consider the following example for community colleges. Assume the work-based component amounts to 75% of the apprentice's learning program and the school-based courses are only 25% of the normal load for students without an apprenticeship. By allowing training providers to keep more than 25% of a standard full-time-equivalent cost provided by federal, state, and local governments in return for providing the classroom component of apprenticeship, the community colleges and other training organizations would have a strong incentive to develop units to stimulate apprenticeships.

Another possibility is to emphasize apprenticeships in the context of existing high school-based career and technical education (CTE) programs. Since high school CTE courses are already financed as an entitlement, the funds to complement work-based learning in apprenticeships would be readily available. Good places to start are career academies—schools within high schools that have an industry or occupational focus—and regional CTE centers. Over 7,000 career academies operate in the United States in

fields ranging from health and finance to travel and construction (Kemple and Willner 2008). Career academies and CTE schools already include classroom-related instruction and sometimes work with employers to develop internships. Because a serious apprenticeship involves learning skills in the workplace at the employer's expense, these school-based programs would be able to reduce the costs of teachers relative to a full-time student. If, for example, a student spent two days per week in a paid apprenticeship, or 40% of time otherwise spent in school, the school should be able to save perhaps 15% to 30% of the costs. Applying these funds to marketing, counseling, and oversight for youth apprenticeship should allow the academy or other school to stimulate employers to provide apprenticeship slots. Success in reaching employers will require talented, business-friendly staff who are well trained in business issues and apprenticeship.

Allowing the use of Pell Grants to pay for at least the classroom portion of a registered apprenticeship program makes perfect sense as well. Currently, a large chunk of Pell Grants pays for occupationally oriented programs at community colleges and for-profit career colleges. The returns on such investments are far lower than the returns to apprenticeship. The Department of Education can already authorize experiments under the federal student aid programs (Olinsky and Ayres 2013), allowing Pell Grants for some students learning high-demand jobs as part of a certificate program. Extending the initiative to support related instruction (normally formal courses) in an apprenticeship could increase apprenticeship slots and reduce the amount the federal government would have to spend to support these individuals in full-time schooling.

The GI Bill already provides housing benefits and subsidizes wages for veterans in apprenticeships. However, funding for colleges and university expenses is far higher than for apprenticeship. Offering half the GI Bill college benefits to employers hiring veterans into an apprenticeship program could be accomplished by amending the law. However, unless the liberalized uses of Pell Grants and GI Bill benefits are linked with an extensive marketing campaign, the take-up by employers is likely to be limited.

Counseling, Screening Prospective Apprentices to Ensure They Are Well Prepared. Apprenticeships typically require apprentices and employers to commit to a long-term, two- to five-year training program. Before making any commitment of this duration, apprentices should have a clear understanding of the occupation they are entering, the production and learning activities they will undertake during the apprenticeship, and the long-term career opportunities that completing the apprenticeship will afford.

In the United States, for those considering entering a youth apprenticeship program, formal counseling does take place in high schools, usually during sophomore year. But typically, US workers enter registered apprenticeships well after high school, when they are in their mid- to late 20s. Although some workers may receive counseling services from American job centers, most learn about apprenticeships informally, having bounced among various occupations and jobs. They learn from media, friends, and families about apprenticeship openings and apprenticeship occupations. The informal knowledge may not be enough for apprentices to appreciate fully what the job, career, and work atmosphere will entail. Still, unlike those going through a degree program before entering a profession, apprentices will learn about the occupation within the first few months of their education and training.

Typically, the screening process brings out information on the test scores in math and verbal as well as work experience and also provides some gauge of how enthusiastic apprentices are when applying to an employer. However, increasing opportunities for apprentices and employers to learn more about each other before an agreement is formalized should be on the agenda for expanding apprenticeships in the United States. Improved systems for matching prospective apprentices with current and future apprenticeships offered by employers could improve this process.¹⁵⁹

Train the Trainers. The quality of trainers is an important element in the success of apprenticeships. That is one reason why several European systems devote considerable time to training and certifying trainers/mentors of apprenticeship. In the late 1990s, the European Centre for the Development of Vocational Education (Cedefop), decided to promote the sharing of best practices for training trainers and other vocational education instructors across 22 national networks.

In Germany, anyone who wishes to serve as a trainer in the apprenticeship system must demonstrate both technical qualifications and appropriate personal attributes. Trainers are skilled workers who have several years of professional experience and have taken a two-week course at a chamber of industry and commerce or chamber of crafts and trades to prepare for the AEVO exam. Trainer aptitude includes the ability to independently plan, conduct, and monitor vocational training, as well as to plan and prepare training programs, to collaborate in the hiring of apprentices, and to conduct and conclude training. Today, some 90,000 people per year take the trainer aptitude examination.

¹⁵⁹For an example of an apprenticeship-matching site, see <https://www.gov.uk/apply-apprenticeship>.

A trainer must be able to successfully engage in numerous tasks: to examine the capacity of the company to offer training in the desired certified trade, to create a company training program on the basis of training regulation geared toward the job-specific work and business processes, to create the necessary conditions and foster a motivating learning environment, to select training methods and materials appropriate to the target group and deploy them in specific situations, to support apprentices with learning difficulties through customized training design and counseling, to prepare apprentices for the final and master worker examination, and finally to bring the training program to successful conclusion.

The United States lacks any formal system for ensuring trainers of apprentices have the requisite skills and personal attributes to perform well.

Research, Evaluation, and Dissemination. An infrastructure for research, evaluation, dissemination, and peer support can play an important role in scaling up and continuously improving the apprenticeship system. Such functions offer clear externalities to workers and employers. The federal government should sponsor the development of a public-private partnership that houses an information clearinghouse, a peer support network, and a research and evaluation program on apprenticeship. Research could be conducted on the effectiveness of apprenticeships in ensuring that workers learn the key occupational, employability, and academic skills; on the short-term and long-term impacts on earnings compared with other approaches to education and training; and on the regulatory aspects of apprenticeship. Also important are topics especially relevant to employers, such as the return to apprenticeship from the employer perspective and the net cost of sponsoring an apprentice after taking account of the apprentice's contribution to production. The evaluations should cover best practices for marketing apprenticeship, incorporating classroom and work-based learning by sector and counseling potential apprentices.

An information clearinghouse can document international experience with apprenticeship, including skill frameworks for apprenticeships used in various countries. Finally, the public-private institute would engage in dissemination about the impacts of apprenticeships and best practices in apprenticeship.

Summing Up

Expanding apprenticeship is a potential game-changer for improving the lives of millions of Americans and preventing further erosion of the middle class. Apprenticeships widen routes to rewarding careers by upgrading skills,

including occupational skills but also math, reading, and employability skills. Taking math, reading, and writing in the context of using these competencies in the workforce will increase the motivation of many workers and the efficacy of the delivery process. Given the ability of workers to learn more, remain well motivated, and notice how to make innovations at the workplace, firms will have an increased incentive to adopt “high road” strategies and make them work. Such an approach may be one of the only ways the firm can attract and sustain workers.

Yet, today funding for the “academic only” approach to skill development dwarfs the very limited amounts available to market and support apprenticeship. Instead of spending well over \$11,000 per year per student on students in community college career programs, why not shift resources toward far more cost-effective apprenticeship programs? Apprenticeship programs yield far higher and more immediate impacts on earnings than community or career college programs yet cost the student and government far less. Community college graduation rates, especially for low-income students, are dismally low. Even after graduating, individuals often have trouble finding a relevant job. For students in postsecondary education, foregone earnings are one of the highest costs and many students incur considerable debt. In contrast, participants in apprenticeships rarely lose earnings and often earn more than if they did not enter an apprenticeship. Rarely must apprentices go into debt while they learn. And apprentices are already connected with an employer and can demonstrate the relevant credentials and work experience demanded by other employers. Another advantage are the net gains flowing to employers from apprenticeship programs.

Structural barriers require some upfront government investments to help build a robust apprenticeship system in the United States. Investments in marketing as well as the development of standards, along with ongoing support for the off-job costs of apprenticeship, are likely to attract large numbers of employers. As more employers adopt apprenticeship strategies successfully, network effects could well take over, with employers learning from each other about the value of apprenticeship. At some juncture, we may see a tipping point when government spending on marketing becomes far less necessary.

It is past time for federal and state governments to make a genuine effort to build an extensive and high-value apprenticeship system. Without such an effort, we are not likely to upgrade skills and jobs but will likely continue to expend vast resources on a college-based, academic-only system that fails millions of students. With such an effort, I believe US employers will follow their counterparts in other countries, create a significant number of apprenticeship slots, and realize gains in recruitment, workforce quality, and

improved productivity. Institutional change of this magnitude is difficult and will take time but will be worthwhile in increasing earnings of workers in middle-skill jobs, widening access to rewarding careers, raising national productivity, enhancing occupational identity, increasing job satisfaction, and expanding the middle class.

References

Bauernschuster, Stefan, Oliver Falck, and Stephan Heblach. 2009. "Training and Innovation." *Journal of Human Capital* 3 (4): 323–53.

Buchanan, John, Catherine Raffaele, Nick Glozier, and Aran Kanagaratnam. 2016. *Beyond Mentoring: Social Support Structures for Young Australian Carpentry Apprentices*. Research report. National Centre for Vocational Education Research.

Cappelli, Peter. 2015. "Skill Gaps, Skill Shortages and Skill Mismatches: Evidence and Arguments for the United States." *ILR Review* 68 (2): 251–90.

Carnevale, Anthony P., Nicole Smith, and Jeff Strohl. 2010. "Help Wanted: Projections of Jobs and Education Requirements through 2018." Center on Education and the Workforce, Georgetown University, pp. 1–5. <http://cew.georgetown.edu/jobs2018>.

Cedefop. 2011. *Vocational Education and Training Is Good for You: The Social Benefits of VET for Individuals*. Research Paper 17. Luxembourg: Publications Office of the European Union.

Cellini, Stephanie. 2012. "For-Profit Higher Education: An Assessment of Costs and Benefits." *National Tax Journal* 65 (1): 153–79.

Clark, Damon, and René Fahr. 2001. "The Promise of Workplace Training for Non-College-Bound Youth: Theory and Evidence from German Apprenticeship." IZA Discussion Paper 378. Bonn, Germany: Institute of Labor Economics. http://www.iza.org/en/webcontent/publications/papers/viewAbstract?dp_id=378.

Deloitte. 2011. *Boiling Point? The Skills Gap in U.S. Manufacturing*. Manufacturing Institute. <http://www.themanufacturinginstitute.org/~/media/A07730B2A798437D98501E798C2E13AA.ashx>.

Fersterer, Josef, Jörn-Steffen Pischke, and Rudolf Winter-Ebmer. 2008. "Returns to Apprenticeship Training in Austria: Evidence from Failed Firms." *Scandinavian Journal of Economics* 110 (4): 733–53.

Geel, Regula, and Uschi Backes-Gellner. 2009. "Occupational Mobility within and between Skill Clusters: An Empirical Analysis Based on the Skill-Weights Approach." Swiss Leading House Working Paper 47. Swiss Federal Office for Professional Education and Technology, Zurich. <http://ideas.repec.org/p/iso/educat/0047.html>.

Geel, Regula, Johannes Mure, and Uschi Backes-Gellner. 2011. "Specificity of Occupational Training and Occupational Mobility: An Empirical Study Based on Lazear's Skill-Weights Approach." *Education Economics* 19 (5): 519–35.

Göggel, Kathrin, and Thomas Zwick. 2012. "Heterogeneous Wage Effects of Apprenticeship Training." *Scandinavian Journal of Economics* 114 (3): 756–79.

Goldin, Claudia, and Lawrence F. Katz. 2008. *The Race between Education and Technology*. Cambridge, MA: Harvard University Press.

Halpern, Robert. 2009. *The Means to Grow Up: Reinventing Apprenticeship as a Developmental Support in Adolescence*. New York: Routledge.

Hanushek, Eric, and Ludger Woessmann. 2015. *The Knowledge Capital of Nations: Education and the Economics of Growth*. Cambridge, MA: MIT Press.

Hanushek, Eric A., Guido Schwerdt, Ludger Woessmann, and Lei Zhang. 2017. "General Education, Vocational Education, and Labor-Market Outcomes over the Lifecycle." *Journal of Human Resources* 52 (1): 48–87.

Helper, Susan, Ryan Noonan, Jessica R. Nicholson, and David Langdon. 2016. *The Benefits and Costs of Apprenticeship Programs: A Business Perspective*. US Department of Commerce.

Hoeckel, Kathrin, Simon Field, and W. Norton Grubb. 2009. *A Learning Jobs Review for Switzerland*. OECD Reviews of Vocational Education and Training. Paris: OECD Publishing.

Hoeckel, Kathrin, and Robert Schwartz. 2009. *A Learning for Jobs Review of Germany*. Paris: OECD Publishing.

Hoffman, Nancy. 2011. *Schooling in the Workplace: How Six of the World's Best Vocational Education Systems Prepare Young People for Jobs and Life*. Cambridge, MA: Harvard Education Press.

Hollenbeck, Kevin. 2008. "State Use of Workforce System Net Impact Estimates and Rates of Return." Paper presented at the Association for Public Policy Analysis and Management (APPAM) November 8. Conference, Los Angeles. <http://research.upjohn.org/confpapers/1>.

Kemple, James, with Cynthia Willner. 2008. *Career Academies: Long-Term Impacts on Labor Market Outcomes and Educational Attainment, and Transitions to Adulthood*. New York: Manpower Demonstration Research Corporation.

Lazear, Edward. 2009. "Firm-Specific Human Capital: A Skill-Weights Approach." *Journal of Political Economy* 117: 914–40.

Lerman, Robert. 2010. "Apprenticeship in the United States: Patterns of Governance and Recent Developments." In *Rediscovering Apprenticeship. Research Findings of the International Network on Innovative Apprenticeship (INAP)*. Erica Smith and Felix Rauner (Eds.). Springer-Verlag.

———. 2014. "Do Firms Benefit from Apprenticeship Investments?" IZA World of Labor. <http://wol.iza.org/articles/do-firms-benefit-from-apprenticeship-investments-1.pdf>.

Lerman, Robert, Lauren Eyster, and Kate Chambers. 2009. *The Benefits and Challenges of Registered Apprenticeship: The Sponsors' Perspective*. Washington, DC: US Department of Labor, Employment and Training Administration. http://www.urban.org/UploadedPDF/411907_registered_apprenticeship.pdf.

Muhlemann, Samuel, and Stefan C. Wolter. 2014. "Return on Investment of Apprenticeship Systems for Enterprises: Evidence from Cost-Benefit Analyses." *IZA Journal of Labor Policy*. <http://www.izajolp.com/content/3/1/25>.

OECD. 2010. *Off to a Good Start: Jobs for Youth*. Paris: OECD Publishing.

Olinsky, Ben, and Sarah Ayres. 2013. *Training for Success: A Policy to Expand Apprenticeships in the United States*. Washington, DC: Center for American Progress.

Osterman, Paul, and Andrew Weaver. 2014. "Skills and Skill Gaps in Manufacturing." In *Production In the Innovation Economy*, edited by Richard Locke and Rachel Wellhausen, 17–59. Cambridge, MA: MIT Press.

Reed, Deborah, Albert Liu, Rebecca Kleinman, Annalisa Mastri, Davin Reed, Samina Sattar, and Jessica Ziegler. 2012. *An Effectiveness Assessment and Cost-Benefit Analysis of Registered Apprenticeship in 10 States*. Washington, DC: U.S. Department of Labor, Office of Apprenticeship. Retrieved from http://wdr.doleta.gov/research/FullText_Documents/ETAOP_2012_10.pdf.

Washington State Workforce Training and Education Coordinating Board. 2014. "2014 Workforce Training Results by Program." Olympia, WA. <http://wtb.wa.gov/WorkforceTrainingResults.asp>.

Reviving America's Forgotten Innovation System: Fostering US Growth through Incremental Product and Process Innovation

Dan Breznitz

Professor and Munk Chair of Innovation Studies and the Co-Director of the Innovation Policy Lab, Munk School of the University of Toronto

Peter Cowhey

Dean and Qualcomm Endowed Chair in Communications and Technology Policy, UC San Diego

Introduction

Everyone in Washington agrees on the need to promote “innovation” to ensure economic prosperity. However, high levels of unemployment have eroded confidence in the ability of our innovation system to create and maintain middle-class jobs for Americans. In our view, “innovation” can sustain employment and competitiveness, but only if we redefine our understanding of innovation and the policies supporting innovation.

America still leads the world in creating new products and services involving game-changing technology that generates enormous wealth and value. This is novel-product innovation, where the creator conjures an entirely new technology or product. It could be a new category of product, such as the first word processor or protease inhibitors, or a new service, such as Facebook. This kind of innovation is knowledge intensive and adds a lot of value. While America must periodically refresh the ecosystem supporting this type of innovation to respond to global competition, it remains by far the best nation in the world at it. The success of leading US firms such as Apple, Google, Genzyme, and Qualcomm shows that these capabilities remain strong. However, novel-product innovation is only half the story of innovation. And America has, at considerable cost to employment, neglected the other half.

The second type of innovation is incremental and process (I&P) innovation. I&P innovation consists of improvements in the ways that goods or services are designed, produced, distributed, and serviced. As the economist Joseph Schumpeter observed, I&P innovation has the greatest impact on

economic growth.¹⁶⁰ Inventing the internal combustion engine did not change society because there were no preexisting devices to put such an engine in. Instead, the wave of innovations that improved and applied this innovation throughout the economy is what changed society. Some industries, including cars and personal computers, are less defined by rapid product innovation than by continuous process improvements, which alter cost and performance capabilities.¹⁶¹ Germany's manufacturing success stems from this latter kind of innovation.

Taking an idea from concept to marketed product requires an array of incremental innovations in the product (such as continual improvements in automobile transmissions) and the production process. Such innovations are often made on the factory floor by workers with intimate knowledge of their product and process. For example, at Inspur, a Chinese high-end computing company, a regular line worker devised a way—now patented—to control static buildup on the production line using plain tap water. This innovation is essential for the production of sensitive electronics. Production innovations improve firm performance and ensure competitiveness, even in the face of rising costs. Once a core American strength, I&P innovation is hampered by our current US system.

The globalization of design, manufacturing, and distribution requires a new approach to I&P innovation if the United States is to avoid losing jobs and production capabilities needed for our competitiveness. Emphasizing only novel-product innovation has two problems. First, novel-product innovation does not, by itself, generate as many jobs because developmental work and production are increasingly offshore from the beginning. Second, the divorce of novel-product innovation from I&P innovation decreases the rate and quantity of novel-product innovation and erodes overall competitive capacity. The United States must address both challenges when considering policies for reinvigorating growth.

The fact that the United States has a problem in the manufacturing sector is made clear by **Figure 1**, **Figure 2**, and **Figure 3**. Figure 1 shows that the number of employees engaged in manufacturing has declined steadily since 1990, even as the US population has grown considerably. Figure 2 shows the declining percentage of the total workforce that is engaged in manufacturing;

¹⁶⁰J. A. Schumpeter, *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle* (Cambridge, MA: Harvard University Press, 1961; originally published 1934).

¹⁶¹Continuity in fundamental end-products does not mean that innovation is missing. While major brands may seem more or less constant over time, all continually innovate in the underlying process technology.

Figure 1. US Manufacturing Employment

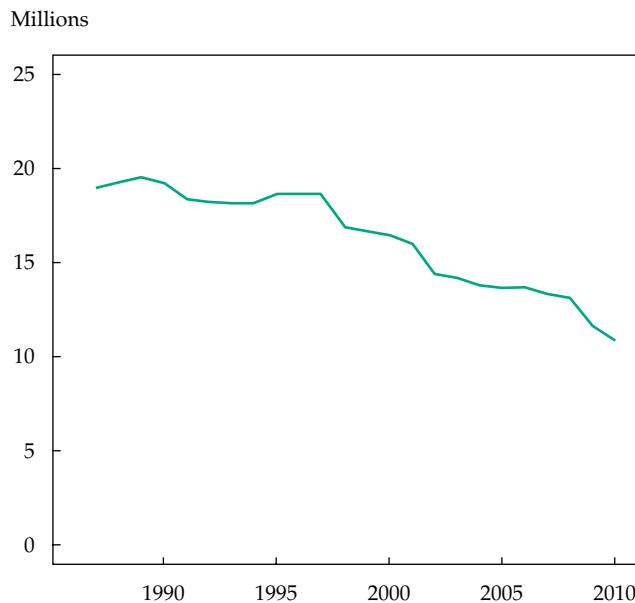


Figure 2. Manufacturing Employment, Percentage of Total US Employment

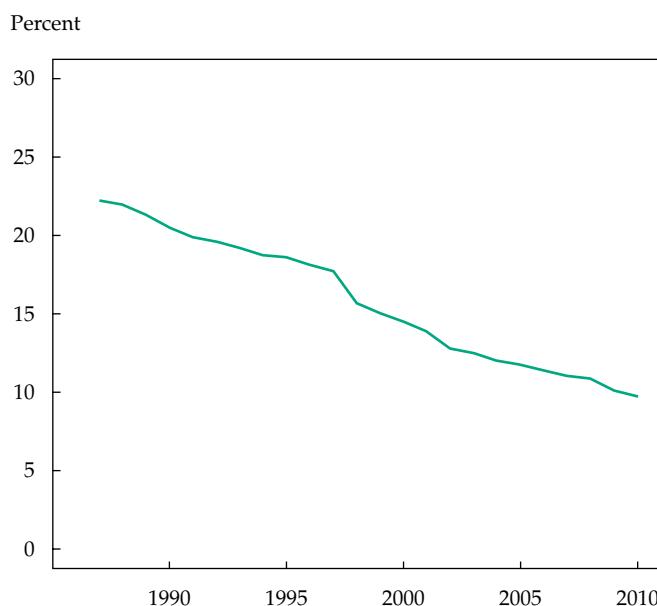
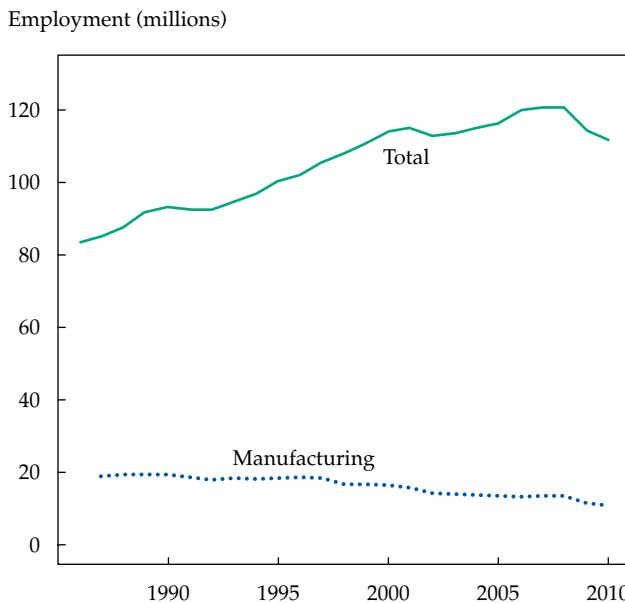


Figure 3. US Employment, Manufacturing Sector, and Total

this decline is steeper than the head count in Figure 1 precisely because of the population growth just mentioned. Figure 3 ties the information in the previous two figures together, showing the data in both absolute form and as a percentage of total employment.

A strategy that bolsters I&P innovation in the United States must recognize the growing interdependence of services and manufacturing in the new global landscape. This paper will refer primarily to manufacturing because that is where the immediate challenge for jobs lies, but the service and information technology (e.g., software and information-enabled service) content of manufactured products is perhaps the fastest growing segment of added value in manufacturing. And service packages, combined with manufactured products, are the backbone of many higher-end products in which the United States should remain competitive. So, when we speak of manufacturing, we are also referring to the embedded service and information base on which manufacturing increasingly depends.

Even if the United States does everything right to foster I&P innovation, a substantial amount of production will never be located in the United States. However, because of structural flaws in our innovation system, production in America is less sophisticated and good jobs fewer in number than is possible. Although America's two innovation systems are in very different health,

they both rest on four fundamental building blocks that address both market mechanisms and the building of social capital. These four building blocks are

1. Shared production assets: Firms need to fund and use assets held in common.
2. Effective innovation network structures: New structures are necessary as markets, contracts, and firms no longer provide adequate “glue” for linking together I&P innovators.
3. Flexible business models: Restructuring the traditional definitions of supply-and-demand functions in markets is often as important as an innovative product.
4. Specialized financial institutions: Risk assessment capacity and lending/investment models appropriate to different types of innovation are necessary.

Based on a study we conducted with the Connect Innovation Institute of San Diego, this paper explains the role of these building blocks and indicates how to strengthen them for I&P innovation.¹⁶² Section II analyzes the different forms and economic impacts of innovation and explains how the current model used in setting innovation policy differs from the vertically integrated model that dominated through the 1960s. Section III analyzes the weaknesses of the conventional model for I&P innovation. Section IV focuses on the four building blocks of I&P innovation and shows how policy could revitalize American I&P innovation. Section V concludes by suggesting specific solutions.

I. The Changing American Models for Innovation

Societies and economies change over time, and so has the American innovation system since 1945. We briefly review the consequences of the change

¹⁶²The Connect Innovation Institute of San Diego sponsored research to develop and publish white papers on production, innovation policy, and competitiveness in the global economy. This paper builds on the insights of earlier white papers, including D. Breznitz and P. Cowhey, “America’s Two Systems of Innovation: Recommendations for Policy Changes to Support Innovation, Production and Job Creation” (Connect Innovation Institute White Paper: Project on Production Innovation, 2012); E. Fuchs, “The Impact of Manufacturing Offshore on Technology Competitiveness: Implications for U.S. Policy” (Connect Innovation Institute White Paper: Project on Production Innovation, 2012); E. B. Reynolds, “Technology, Policy and Product Life Cycle: The Evolving Geography of Biomanufacturing” (Connect Innovation Institute White Paper: Project on Production Innovation, 2012); J. Whitford, “Network Failures and Innovation in the New Old Economy” (Connect Innovation Institute White Paper: Project on Production Innovation, 2012).

in our innovation system. Our key point is this: The renewal in the innovation system that reinvigorated the novel-product innovation sector in the last 30 years had the collateral consequence of stranding the I&P system without adequate policy support.

Before our current model of innovation policy emerged, US research spending and technology production after 1945 involved large enterprises conducting both novel-product and I&P innovation in-house. Many enterprises worked in both defense and civilian markets. Firms had enormous financial and human capital resources, including pools of “patient capital” that could be invested without expecting rapid returns. Firms were vertically integrated, a structure that encouraged networking among specialist groups in different phases of design and production. For example, AT&T’s Bell Labs designed its facilities with long halls, forcing researchers to walk past other labs, opening chances for conversations and unforeseen collaborations. Firms internalized many of the financial risk management functions for innovation.

This model of innovation has receded sharply. Since the corporate restructurings of the 1970s and 1980s that were designed to respond to Japan’s economic challenge, firms have focused on “core competencies,” leading to a new landscape where a great deal of innovation is driven by entrants focused on specific stages of production. This significant change enabled the development of new arrangements for financing and interconnecting the newly fragmented ecosystem of innovation.

Our current model has thus favored startup-driven, novel-product innovation. This model focuses on interaction at the national and regional levels through geographic technology clusters. In this model, anchor universities are critical to both knowledge creation and the development of human capital. Essential to this model are new laws and regulations that allowed and incentivized the creation of financial vehicles, such as venture capital (VC) funds, and the opening of markets (such as NASDAQ) that allow realization of financial gains within a short timespan. Technology clusters also strive to develop an ecosystem of professional support services for new technology-based firms (e.g., law and accounting firms). The great success of this model and the immense financial gains that accrued to founders and financiers make it the focus of policy discussions. Fostering technology clusters has become America’s policy mantra. The existence and success of regional anchors makes the conventional model, which involves supportive federal policy, politically viable in both blue and red states.

Successful novel-product innovation clusters address market failures and social/informational networking. When researchers are not part of internal corporate innovation systems, other institutions and practices for obtaining financing

arise. Among these are courting venture and angel capitalists attuned to the region and promoting incubators that lower costs and identify prospects for early investors. Significantly, common economic and research assets for firms (e.g., mass spectrometers) are often created, frequently in anchor universities. Clusters provide networking organizations for technology specialists; experts circulate freely and share knowledge through a variety of events.

Cluster leaders routinely acknowledge that people are the most important asset for novel-product innovation industries. Clusters provide a substitute for the human networks that arise in vertically integrated firms. Research in economic sociology and geography underscores the idea that strong social networks, abetted by formal and informal institutions, are essential to the circulation of knowledge and people, and thus the building of trust, that makes for successful clusters.

The divergence in the paths of the California and Massachusetts information industries has been attributed to the difference in their social networks. Underlying the regional clusters are national policies that fund the basic and applied R&D needed to induce the training of researchers and engineers, protect intellectual property central to novel technologies, and enforce rules of competition that keep markets open to newcomers.

This model is well suited to novel-product innovation in the United States. It will largely continue to be, thanks to four strengths:

1. Basic R&D and research universities—not only does the United States remain the dominant science research power, it has a great regional spread of specialized strengths.
2. Ease of successfully commercializing new ideas early on, a result of America's strong rule of law, relatively easy market entry, and availability and quality of entrepreneurs and professional services.
3. A sophisticated user base that co-invents innovations and suggests flexible value and business propositions—where and how money is made can be surprising and innovative.
4. The world's best system for mobilizing financial resources to turn ideas into businesses. Startups can quickly attract \$200 to \$400 million without being seen as an anomaly.

Nonetheless, although it is not the focus of this paper, we note the need to update this model in light of rising international competition and the global fragmentation of production and innovation. In particular, financial incentives, including the decline in US government funding for basic and applied research, are undermining the US advantage.

II. The Faltering of Incremental Product and Process Innovation

Although the current system of startup-driven clusters shored up American leadership in novel-product innovation, it inadvertently weakened US leadership in I&P innovation. This is troubling for two reasons. Setting the wrong incentives for I&P innovation weakens the benefits of novel-product innovation for job creation. And, as Erica Fuchs's research has shown, these failures weaken the long-term sustainability of novel-product innovation.¹⁶³

To illustrate the challenges, consider the role of manufacturing in the United States. Contrary to popular belief, as recently as 2009, the United States was still the world's largest manufacturer with about \$1.6 trillion in output, 21% of the global total (see **Figure 4**).

Although the United States began offshoring manufacturing in the 1970s, the level of domestic manufacturing output has continued to rise steadily, except during recessions. Even with real declines in employment since 2000, manufacturing still supports about one in six private-sector jobs (18.6 million, of which 12 million are directly in manufacturing).

Despite these statistics, the larger picture for manufacturing is not healthy. Manufacturing employment is declining for reasons other than rising productivity or uncompetitive wages. The distribution of manufacturing jobs is also becoming increasingly uneven. While the Midwest was once the largest locus of manufacturing, states such as Arkansas, Mississippi, and Alabama are now more heavily industrialized than the former bastions of industry (see **Figure 5**).

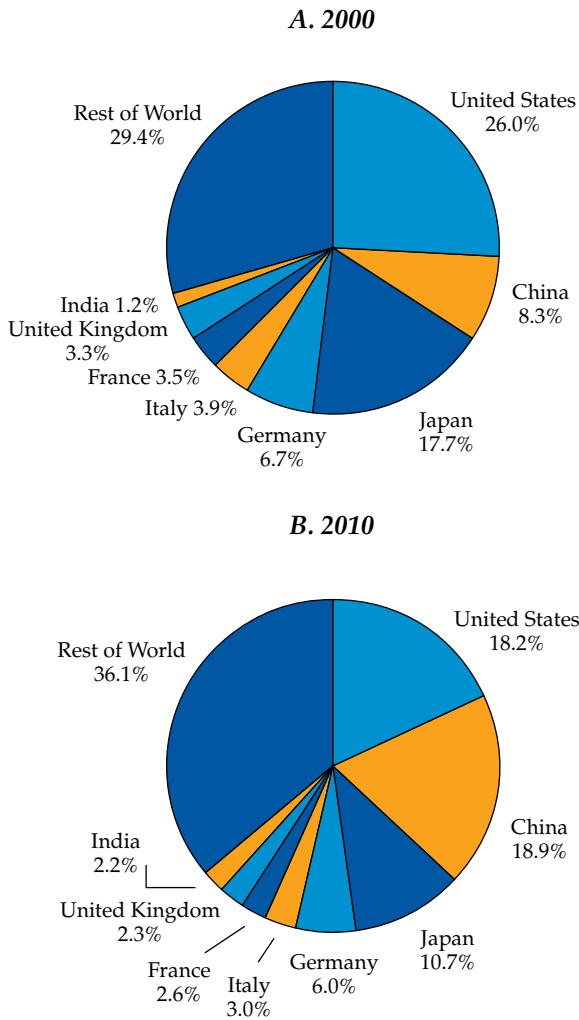
Nationwide, however, the trend remains negative. Germany and Japan have both high wages and high productivity, yet manufacturing remains a larger part of their economy than it is of ours (roughly 20% of GDP versus 12.8% in the United States).¹⁶⁴ See **Figure 6**.

The decline in manufacturing employment, and the tilt toward foreign production by American companies instead of the construction of new domestic facilities has roots elsewhere in the American I&P innovation system.

¹⁶³E. Fuchs, "The Impact of Manufacturing Offshore on Technology Competitiveness: Implications for U.S. Policy" (Connect Innovation Institute White Paper: Project on Production Innovation, 2012).

¹⁶⁴J. Whitford, "Network Failures and Innovation in the New Old Economy" (Connect Innovation Institute White Paper: Project on Production Innovation, 2012).

Figure 4. Shares of World Manufacturing Output, 2000 and 2010

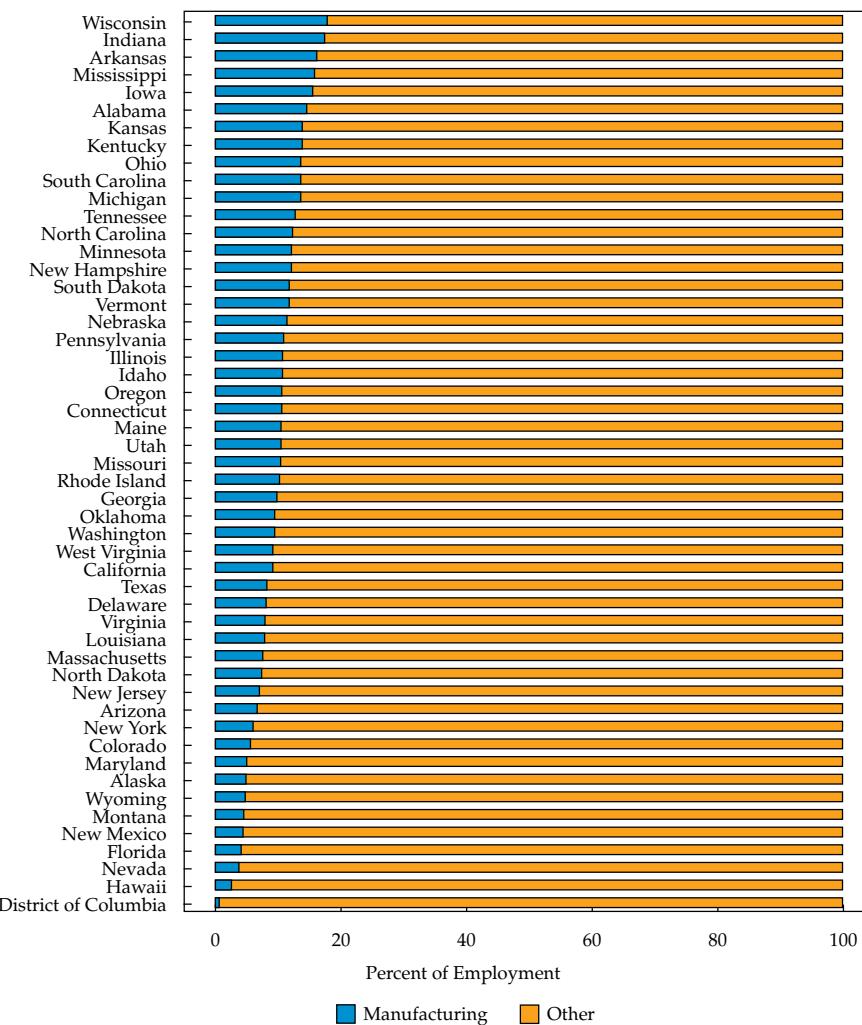


Source: A. Mellows-Facer and L. Maer, "International Comparisons of Manufacturing Output," Standard Note: SN/EP/5809 (London: UK Parliament, 2012).

III. Underlying Stresses Eroding the I&P Model

The weakness in our current I&P innovation system has two major causes.

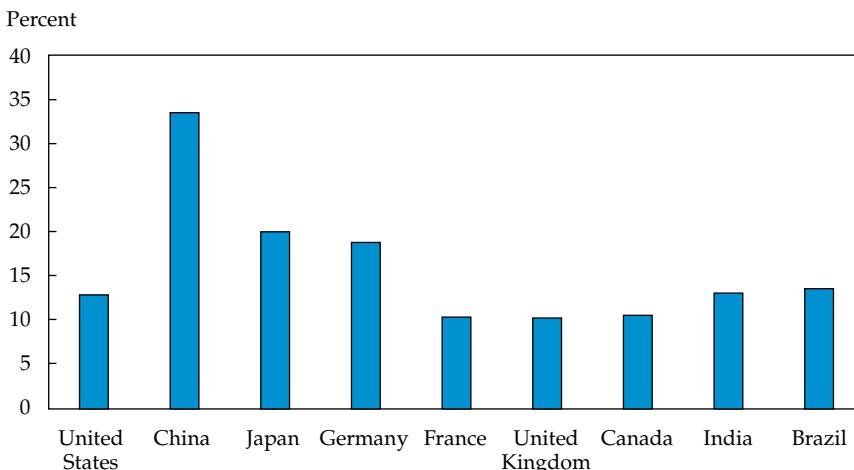
First, the production of products and services has significantly changed. We live in a world of fragmented production that has eroded the networking necessary for successful I&P innovation. Companies that are part of production

Figure 5. Manufacturing and Other Employment, by State

networks specialize in a narrow set of activities, whether high-level R&D, design, manufacturing, or assembly. If Apple once imagined, designed, coded, and assembled computers in its own factories, it now distributes these activities among a multitude of companies in the United States and (especially) Asia. Fragmentation allows companies to profitably specialize in narrow activities.

The innovation and financial needs of companies in different stages of production vary significantly. Fragmentation requires new ways to collaborate among companies and across different modes of operation. We now live

Figure 6. Manufacturing as a Share of GDP (2010)



Source: UNStats, “National Accounts Main Aggregates Database,” United Nations Statistics Division, New York, 2011. Available online at <http://unstats.un.org/unsd/snaama/dnllist.asp>.

in a world of networks *between* companies and organizations, not within them. Innovation ecosystems require policies to solve semipublic good supply problems, such as creating shared production facilities, training, and codevelopment of nonpatentable innovation that are beyond the capabilities of individual firms. Accordingly, the need to fix network failures, not solely market failures, is becoming critical.

Second, the US financial system discourages large capital investment in domestic production or production innovation. Disincentives include the focus on startups and novel-product innovation, the financial constraints under which US public companies operate, and the financial vehicles open to private companies. These factors make it difficult to take risk by making long-term capital investments. More and more jobs, even in the most advanced niches of the high-tech industry, are not being created domestically because of these constraints and the emergence of foreign-based contract manufacturers. American production is not offshored, it is born foreign.

These structural problems manifest themselves through weaknesses in all four of the building blocks of I&P innovation, especially for small and medium-sized enterprises (SMEs) that are the core of sustainable I&P innovation systems:

1. SMEs suffer from inadequate shared assets to complement firm-specific assets. The changing mix of skills necessary for production and

incremental product innovation, especially under fragmented production, falls outside of the traditional core.

2. SMEs lack strong networking institutions to foster knowledge and social capital sharing. I&P innovation requires more than the circulation of smart people such as is fostered in conventional clusters. Context and craft-oriented innovations require structured forms of networking within and among firms.
3. The US legal and regulatory systems do not block business model innovation, but many SMEs depend on subcontracted work; the rules and practices for subcontracting in the United States do not foster innovation.
4. The United States lacks financial institutions with the business models and risk assessment capabilities to analyze and invest in productive capacity by firms.

For our purposes, findings in three studies (Helper and Kuan; Reynolds; and Helper, Krueger, and Wial) are convenient starting points for a discussion of the possibilities and weaknesses—and consequences of weakness—of I&P innovation in the United States.¹⁶⁵

1. High-wage, high value-added production is essential for strengthening the manufacturing employment base; biotech and advanced hardware production in information and communication technology (hereafter “ICT”) are exemplars of possibilities for expanded domestic production.
2. In industries where supply chain issues (time and shipping costs) may allow for increasing lower value-added production, a major obstacle is that only a minority of US suppliers engage in significant I&P innovation. Suppliers do not have the institutional structure to support these activities, thereby weakening domestic production.
3. The composition of added value in manufacturing is shifting. The share of information value-added (software and computing services) in manufacturing is rising, making these inputs more central to I&P innovation. This

¹⁶⁵S. Helper and J. Kuan, “Overcoming Collective Action Problems in the Automotive Supply Chain” (Connect Innovation Institute White Paper: Project on Production Innovation, 2012); S. Helper, T. Krueger, and H. Wial, *Why Does Manufacturing Matter? Which Manufacturing Matters? A Policy Framework* (Washington, DC: Brookings Institution, 2012); E. B. Reynolds, “Technology, Policy and Product Life Cycle: The Evolving Geography of Biomanufacturing” (Connect Innovation Institute White Paper: Project on Production Innovation, 2012).

means the skills for success in manufacturing are shifting. Increasingly, manufacturing requires IT savvy, an area where training and human resource development could bear fruit.

Existing research highlights several policy omissions that hinder I&P innovation. These in turn harm US employment prospects. These omissions may also adversely affect novel-product innovation in some fields.

To illustrate, Erica Fuchs's research finds weaknesses in I&P innovation and domestic manufacturing that erode US firms' novel-product innovation capabilities.¹⁶⁶ Multiple problems hinder the process innovation crucial for domestic production of both the advanced optoelectronics and the automotive industries. Fuchs demonstrates how, at least initially, it is more profitable to produce goods using older technologies in markets where manufacturing is cheaper than to produce more innovative goods using the latest technology in the United States. Because a US production strategy would be both more technically challenging and capital consuming, all publicly traded companies shunned that option for production, and startup financing cannot cover the capital needs for production.

In optoelectronics this obstacle led all publicly traded US firms to offshore production, leaving only private startups to advance new production technologies. Larger firms without productive capabilities do not innovate in these areas, thereby ceding potential new areas for US advantage and advancement. Fuchs found that companies that offshored their production innovate less. This finding suggests that the loss of production activities can rapidly lead to a sharp reduction in a firm's innovation capacities.

These findings also hold for new material technologies for car production. This is alarming for long-term American prosperity, since decisions to produce using less innovative, but more easily offshored, technology lead to a technology trajectory that works *against* US leadership: The most advanced production technologies, where the United States still has a sustained edge, never reach the market.

Equally troubling is that the supply base of middle value-added firms in the complex US supply chain has significant problems realizing its innovation potential. Many leading manufacturing companies, such as car companies, have delegated more and more critical manufacturing, design, and innovation responsibilities for their final product to their supplier base. This means that in order to thrive, leading US manufacturers rely on the innovation vibrancy

¹⁶⁶E. Fuchs, "The Impact of Manufacturing Offshore on Technology Competitiveness: Implications for U.S. Policy" (Connect Innovation Institute White Paper: Project on Production Innovation, 2012).

of their supplier network. But these leaders fail to nurture innovation in the supply chain.

For example, Helper and Kuan's (2012) survey of innovation in automotive supply found that American suppliers view American car companies as less trustworthy partners than their foreign competitors because US car makers provide comparatively little "feeding" of the innovation function in their supply base.¹⁶⁷ As a result, only a minority of firms routinize innovation. This practice damages the industry's competitiveness.

More generally, the current ways of pooling expertise among SMEs are largely ineffective. First, to excel in I&P innovation, firms must have workers with specific skills. Today, firms have too little stability to invest heavily in individual skills, and because of free-riding risks, they fail to sponsor inter-firm training. Similarly, ICT applications are changing production processes and opening ways to more tailored incremental product innovations. Yet, ICT specialists and the new generation of design specialists who can apply ICT products or processes are *not* typically employed by traditional producers, because of their focus on core competencies. The challenge is to find ways of speeding up experimentation and diffusion of ICT innovations across SMEs (David 1995).¹⁶⁸

An even bigger challenge for American production is the limited availability of shared assets for production. Reynolds illustrates the adverse impact of these problems on onshore US production in the case of the biopharma industry.¹⁶⁹ As a technology matures, the industry can modularize production and reduce the risks associated with production. Coupled with regulatory oversight slowly converging across national boundaries, companies have begun to shop for incentives like favorable tax treatment. In addition, surplus biopharma production capacity has required consolidation of facilities. At the same time, the rise of foreign contract manufacturing organizations (CMOs) with local government backing has created sophisticated offshore suppliers that reduce the need for onshore US production. This situation has led to a sharp decrease in new production facilities in the United States, while other high-wage economies, such as Ireland, Denmark, and Switzerland, have become production hubs for American companies.

¹⁶⁷S. Helper and J. Kuan, "Overcoming Collective Action Problems in the Automotive Supply Chain" (Connect Innovation Institute White Paper: Project on Production Innovation, 2012).

¹⁶⁸P. David, "The Computer and the Dynamo" (working paper, Center for Economic Policy Research, Stanford University 1995).

¹⁶⁹E. B. Reynolds, "Technology, Policy and Product Life Cycle: The Evolving Geography of Biomanufacturing" (Connect Innovation Institute White Paper: Project on Production Innovation, 2012).

There are parallels to the US electronics industry. The question is whether pharmaceutical CMOs are more like the “rote” CMOs in electronics, like Hon Hai/Foxconn, or more like the “creative” CMOs that contribute significantly to I&P innovation, like TSMC. Rote CMO activities will be difficult to onshore because of labor cost sensitivity. However, if pharmaceutical CMOs are the “creative” kind, we should ask how they might emerge in the United States. Creative CMOs could reinforce US leadership in the biotech industry and maximize job creation.

A shortage of financing and the inability of stock price-sensitive firms to invest in the most modern production capabilities mean that domestic CMOs are a valuable option. For example, if US shared production facilities for co-use were available, all the firms in Fuchs’s study would have used them and developed products using the latest technology. Having a CMO would make it possible to keep manufacturing and push for more sophisticated production innovation.¹⁷⁰

Both Fuchs and Reynolds show how one of the issues facing US production is the mismatch between the financial model driving creation of research-intensive companies and technology-based publicly listed companies, and financing conducive to long-term capital investment for domestic production. The venture capital system lacks the patience for production; and the stock market cannot easily price long-term investments, so it punishes them.

William Lazonick’s work delineates the pressures against production created by the “financialization” of US corporations.¹⁷¹ In theory, the “financialization model” focuses on disciplining management to emphasize return to shareholders. Emphasizing companies’ core competencies is one path to maximizing shareholder return. Firms should focus on areas where they have sustainable advantages and renovate cost and product structures constantly. Financial markets “monitor” publicly traded firms by emphasizing quarterly

¹⁷⁰E. Fuchs, “The Impact of Manufacturing Offshore on Technology Competitiveness: Implications for U.S. Policy” (Connect Innovation Institute White Paper: Project on Production Innovation, 2012).

¹⁷¹W. Lazonick, *Sustainable Prosperity in the New Economy: Business Organization and High-Technology Employment in the United States* (Kalamazoo, MI: Upjohn Institute for Employment Research, 2009); W. Lazonick, “The New Economy Business Model and the Crisis of US Capitalism,” *Capitalism and Society* 4, no. 2 (2009); W. Lazonick, “The Explosion of Executive Pay and the Erosion of American Prosperity,” *Enterprises et histoire* 57 (2010): 141–164; W. Lazonick, “The Innovative Enterprise and the Developmental State: Towards an Economics of ‘Organizational Success,’” (paper prepared for Institute for New Economic Thinking 2011 Conference, Bretton Woods, NH, 8–11 April 2011); W. Lazonick and M. Mazzucato, “The Risk-Reward Nexus in the Innovation-Inequality Relationship: Who Takes the Risks? Who Gets the Rewards?” (working paper, 2012); W. Lazonick and M. O’Sullivan, eds., *Corporate Governance and Sustainable Prosperity* (New York: Palgrave, 2002).

financial returns and reward investment in only the highest return alternatives. While not impossible (as shown by Intel), massive spending on capital-intensive investments with long payback periods has a steep threshold for approval under this approach.

In addition, the metrics for financial performance and return to shareholders, together with new financial regulations about how to account for minority investments in other firms, further tilt incentives toward the management of quarterly earnings through expanded use of stock buybacks and other devices. Boards reinforce this pattern by providing managerial incentives tightly woven around stock price. But by strongly incentivizing managers to pay attention to stock price, the financial system reinforces short-term financial engineering over long-term strategic production investments, a consequence best avoided.

Moreover, the VC model is not well suited for financing expansion into large-scale production and I&P innovation because it is based on a high-risk, high-return, limited-time model. The goal is to invest in companies that offer financial returns of 100 or more per dollar invested within five years.

Unlike the VC model, most production-innovation focused SMEs aim to increase the profitability of established revenue streams, usually in the low double digits when expressed as an annual rate of return, and they have no wish to sell ownership. Although the VC industry has deep knowledge of novel-product innovations and new enterprise formation, it has little ability to judge the value of I&P investments. Judging I&P requires investors with deep knowledge of that aspect of industry and technology. Industrial banks once possessed such knowledge, but they are no longer part of the US financial landscape. As a result, for both large firms seeking new productive capacity and SMEs seeking to upgrade, the American financial system militates against increasing domestic production assets.

In the next section we suggest that such shared assets as skills training in manufacturing should be treated differently. Since changes in global production tilted the rationale for private actors' investments, we now should no longer assume that the production agents, by themselves, will innovate or provide training and skills development. We borrow from prior American experience and the experiments of other countries to suggest an alternative path forward.

IV. Moving toward Solutions for I&P Innovation

Constraints and Priorities for Policy Solutions. Our approach to reviving up the innovation system in the United States assumes two constraints

on the available policy choices. The first constraint is our divided system of political power and the key role of federalism. Even if someone thought it to be wise, a centralized top-down industrial policy would be impossible for the United States. The importance of competition in US economic policy reflects America's fragmentation of power, as dominance by any market player surely upsets firms in other regions of the country. The second constraint is that the fragmentation of the US system of governance reinforces the natural tendency to have a wide variety of performance capabilities among agencies and administrative domains. The ability to execute policies to nurture innovation varies substantially from region to region.

In light of these constraints, we focus on changes in policy for enhancing I&P innovation that are consistent with strong interfirm market competition. We seek to expand the range of financial tools and organizational/business models available to firms and to increase capabilities through coordination of actors by a variety of mechanisms. We do not advocate strong planning or directing roles for the federal or state governments. We emphasize an I&P focus rather than a manufacturing focus built on specific manufacturing processes. Our approach avoids the tired fights over manufacturing versus services.

Finally, we argue that the right approach is for the federal government to respond to I&P priorities developed from regional discussions because they are best able to build on the results of a fruitful exercise in federalism where local, state, and national authorities have cooperated in the past. A reformulated strategy built on regional anchors with federal support opens the way for winning necessary bipartisan support.

Although we do not dwell on issues of trade policy that influence global competitive conditions for US firms, it is an important complement to the realignment of domestic policies. Of course, I&P innovation is essential for both large and small firms, but SMEs are at the heart of the supply base that I&P innovation can especially reinvigorate. Economic evidence shows that entering world markets is an accelerator of revenue growth, profits, and job growth for firms. Therefore, the new emphasis in American trade policy on trade facilitation for smaller firms is a smart way of accelerating beneficial changes through trade.

Our discussion of the failures in the building blocks for I&P innovation suggest two major themes for reform and policy making:

First, the United States should move beyond regional clusters to platforms. As noted earlier, clusters allow the industrial community to excel in novel-product innovation. Some success comes by addressing market challenges, such as increasing financial options for innovation or promoting

shared use of expensive scientific infrastructure for SMEs. Clusters built social network institutions that promote trust, information transfer, and joint problem solving. Clusters organize regions into networking and information systems to enhance density of interconnection, flow of human capital, and knowledge transfer. Knowledge transfer includes networking and facilitating business contacts. These activities make it easier to match ideas to financiers and support services and to provide specific needed information. This model, established in San Diego, Palo Alto, and Boston, is highly effective in fostering the creation of novel-product, innovation-based enterprises. It is highly capable of providing financing to develop novel-product, innovation-based firms. It fails, however, in I&P innovation.

For success in I&P innovation, clusters must become regional platforms. In entrepreneurial startup-based clusters, churn among employees and the constant start, failure, restart, and boom of companies enables human resource sharing and technology dissemination, but churn is not enough for I&P innovation. Innovation and production are now done via semipublic goods provisions, such as shared production assets. These are critical in more traditional industries, such as metallurgy or automobiles, where there is constant need to spur and diffuse innovation across many SMEs.

Thus, unlike novel-product innovators, the ability of I&P innovators to appropriate gains is limited. Shared assets help to solve this problem of necessary but economically underincentivized knowledge dissemination. Shared assets can be truly shared facilities (owned jointly by multiple companies under a variety of contractual forms), or private, for-profit organizations focused on production for other companies. Pooling productive capacity provides a focal point for development of new tacit production knowledge and enables firms to comfortably share their know-how since the producer is not in direct competition. Better known as CMOs, such firms do not create their own products but rather emphasize I&P innovations as a production base for other firms. A successful example is Hospira, the leading producer of injectable pharmaceuticals, which began when Abbott Laboratories spun off its production division.

Second, the failure on all four building blocks suggests the need to find solutions to network failures for I&P innovation. Having common production assets is not a panacea; some SMEs will and should continue to produce in-house. Sharing solutions within a supply network or introducing missing skills to a network will be critical. This requires bridging the knowledge silos of different industries and technologies, a very difficult task.

For example, it is becoming apparent that in the United States, just as in Israel, many production companies, especially in traditional industries, lack

crucial new skills in areas such as ICT, where the greatest promise for production innovation and improved productivity lie. The United States must also find an appropriate industrial research system model. The German, Taiwan, and Korean markets are possible models, although their very different industrial histories and political economic systems make rote imitation impossible.

Specific Solutions: Regional Platforms. The aim of regional platforms is to solve the specific problems of supplying semipublic goods and addressing network failures as discussed above: the need for joint production assets, the routinization and dissemination of innovation in SMEs, the provision of necessary human talent, and the ability to provide investments for production. Regional platforms do so by *creating common regional assets shared by all companies in an industry*. Platforms alleviate collective-action problems, supply missing critical resources, change firms' risk and profitability calculations before they offshore production, and enhance conversion of ideas on I&P innovation within the United States. Successful platforms serve as seeds of new production-focused American companies that can successfully compete with the best foreign CMOs.

Promoting these platforms specifically acknowledges that I&P innovation, critical to production capabilities, may require more than the circulation of people and a common research base. Instead, as the biopharma case suggests, we should create common assets vital to production—through CMOs and joint assets such as testing and certification centers. How best to do this cannot be answered based on existing research.

Moreover, it is not the place of the federal government to dictate exactly how regional platforms should create common assets. We welcome variations depending on particular industrial structures. Regional experimentation will help ensure that the best models emerge.

This experimentation may lead to significant upheaval in regional, or even the national, economy because even I&P innovation sometimes requires major shifts in business models that upset expectations about how markets work. Henry Ford's Model T melded process and business model innovation. It combined mass production with the mold-breaking business model of pricing cars for working households and paying employees enough to consume their own products. Similarly, Apple's iPod was a breakthrough success through its slick hardware/software/service combination of the iPod and music store. The business model inverted conventional wisdom on pricing: charging a premium for hardware while making content cheap, thereby turning songs back into a product that customers were willing to buy. Such large-scale experiments reinforce the merits of a federalist structure for innovation policy.

As part of the process of experimentation, we note three complementary guidelines.

First, carefully consider the record of CMOs overseas. Low-end “rote” contractors, specializing in the cheap fulfillment of comprehensively blueprinted orders, are not an option; America has no comparative advantages in these low-paying, low-skill jobs, nor would American workers relish such employment or working conditions. However, the example of TMSC and the CMO activities of Samsung deserve careful consideration. If labor costs are no longer the controlling factor, as in sophisticated CMOs, what are the obstacles to such organizations in the United States?

Relatedly, how can excess capacity be repurposed into CMOs instead of closing when their controlling companies opt out of production? As one executive with whom we had many discussions has pointed out, the surplus capacity in biotech production need not be a deadweight loss. This capacity could be the basis for regional shared facilities that grow into successful CMOs attuned to the specifications of their regional innovation platform. Once productive capacity is pooled as a CMO, the new organization must embrace its role as a producer and seek ever more productive means of manufacturing. Local government policies such as tax incentives may aid in this process.

It is not enough to simply create a new CMO or spinoff a division of a formerly vertically integrated company. Creating shared assets, including training of workers, fosters new business models. In the semiconductors industry, the creation of the “pureplay fabrication facilities” specializing in the production of integrated circuits designed by other companies gave viability to the “fabless” IC design company while creating a new business logic emphasizing scale, quality, and security for manufacturers.

There will be a need to increase the range of financial options for different combinations of risk/reward situations involving CMO innovations. The VC model and the predominant mix of current American finance practices do not fit these situations. Developing regional platforms, such as shared production capacity, requires a wholly new ecosystem.

Second, increase specialized regional training schemes. Lack of skilled manufacturing personnel is often cited as a force against increased manufacturing in the United States. Even where a CMO is created, ensuring its I&P innovative capacity and effective operation requires rapidly developing and disseminating human capital. In North Carolina’s research triangle, unique industry-university collaboration around specialized training of workers for biotech production is a key to the region’s tremendous success. Funding for developing human capital comes from both public and private sources and is administered by the local government in concert with industry and the

university system to identify specific needs and strengths of their region. Such a model could be expanded to include representation from local CMOs to determine the required types of human capital and how best to train or attract it.

Third, encourage new financial options for firms specializing in I&P innovations. One approach might be to create new specialized public-private investment banks. The goal would be to convert relatively inefficient state subsidy streams into more leveraged banking schemes for I&P innovation. A pioneer of this approach is the Connecticut Clean Energy Finance and Investment Authority (CEFIA), which has converted a subsidy fund collected by state utility customers into a public investment bank for clean energy projects. Such banks, if able to overcome politicized funding and risk assessment practices, could be oriented toward regional cluster platforms.

A broader approach regarding finance would include instituting changes in deferral regulation and taxation. The purpose would be to combat current taxation and regulation regime incentives for public companies not to invest in domestic production activities. The best job creation agents in the American economy are private companies. It is worrisome that during the recent great recession, the best American corporations, such as Apple, sat atop the largest piles of cash in the history of corporate America. It behooves federal policy makers to think about ways that changes in taxation, and perhaps matching funds, would tilt companies' calculation about return on investment and risk so that they invest at home and not view this action as against shareholders' interests.

Specific Solutions: Network Solutions. The second policy approach is to create new forms of regional networks specializing in "network solutions" to upgrade capabilities for I&P innovation. There are two significant issues with production innovation. Much of this innovation is not protectable under the current intellectual property regimes, and it is crucial to ensure rapid diffusion and wide sharing of recent innovations across supplier networks. This combination of issues creates a perverse outcome. Lacking the ability to extract returns from their investment, many SMEs at the core of the supply base underinvest in innovation. A solution cannot be solely to strengthen intellectual property rights (or to provide direct subsidies to innovate), since this would aggravate the second problem.

Some solutions for information sharing make problems worse. Studies of I&P innovation indicate that traditional systems of information sharing often do not lead to detailed problem solving; worse, they can lock firms into relatively narrow circles of expertise and transaction contacts.

To upgrade SMEs' capabilities in, say, industries that are not R&D intensive, such as metal-bashing, it is critically important to build bridges across industry segments.

Our first prong of networked solutions is to embrace public funding for an R&D system aimed at supporting the production (of goods and services) that emphasizes the networking benefits of R&D. There is agreement that an individual firm may not capture all the returns from research and may therefore underinvest in this aspect of knowledge creation. But there is disagreement about the size of this disincentive and about the appropriate level of public funding to bolster applied research. This debate over the size of the market failure ignores the proven record of such research efforts, properly defined, in promoting the networking of knowledge and innovation among smaller firms.

The weaknesses of the American I&P innovation system have led some analysts to suggest that the United States should look at how other markets have utilized public research institutes to solve similar issues. The examples include Korean research institutions, the German Fraunhofer institutes network, and Taiwan's ITRI. These public research institutes have similar designs: Their specialized departments (or sub-institutes) focus on particular industrial niches and sets of technologies, develop long-term relationships with industry, and establish a division of labor.

In this division of labor, the institutes concentrate on the core and continuous production of R&D, and they diffuse the results widely to industry, which focuses on final development and implementation of these technologies. In Taiwan, ITRI's efforts are well known for successfully performing applied research within the institute and disseminating the findings, either to existing Taiwanese firms or spinning off the technology and research group as firms. Some propose setting up regional level Fraunhofer-like organizations in the United States.

However, it would be difficult to duplicate institutions that owe much of their success to cultural environments extremely different from those of the United States. A more productive precedent for the United States is our experience with agricultural research. In agriculture, the government assumption was similar to that in low- and mid-tech industries in other countries with industrial research institutions: Farmers cannot be expected to fund or carry out innovation or technology implementation on their own. We should not assume that firms can conduct necessary innovation, or even independently acquire the skills, to continuously excel in the market. Hence, a program for production innovation utilizing the organizational logic of agriculture research in the United States, devised along regional specialization and

sponsored at the state and the federal levels, offers a better fit for the United States.

Creation of regional innovation information systems could be an important tool for an I&P innovation “extension” system. Such systems can resemble the technological roadmaps that proved so useful to coordinating investment and spurring innovation in many domains, such as semiconductors. Other countries use public-private dialogues at the national and regional levels to identify essential technology building blocks for innovations in particular industries. To fit the agricultural extension model, they need to focus on specific regional I&P objectives, such as how to better match improvements in production processes to novel-product innovations underway in a region. The key challenge in these dialogues is to turn an expert group exercise into a broader community discussion to validate and amend the maps to make them widely acceptable to and desired by industry.

These exercises can be powerful complements to public commitments for change. For example, in China, part of the abundance of funding for photovoltaics is the result of a government bank, but an even larger part is the commercial banking system’s conviction that government roadmaps for reducing emissions cannot be met without photovoltaics. Knowing the strong commitment to reducing emissions, banks are willing to invest because they are confident the market will emerge, even if the state does not directly mandate the technology.

The second way to implement “networked solution” systems is to create regional problem-solving teams that draw expertise from different segments of the supply chain. Other countries’ successful experiments with this approach have the following characteristics:

- a) bridging traditional segments within an industry, thereby maximizing networking contacts
- b) bridging traditional industries, and new technologies and the skills needed to operate them, hence infusing them with new knowledge, ideas, and the skills to act upon them
- c) being governed by multistakeholder boards, including government officials, so as to drive responsiveness to new group demands
- d) focusing on solving problems and creating shared capabilities (such as quality testing labs) for the network through engagement of members of many organizations in the network.

In the words of McDermott, Corredoira, and Kruse (2009, p. 1292), such networks can “provide firms with a new scale and scope of diverse services

and foster new learning relationships between firms from previously isolated producer communities.”¹⁷²

The regional base for such organizations plays an important role in their success because they can develop informal transactional mechanisms that are more effective than standard contracts and rules when dealing with the uncertainties characterizing innovation. They also provide important feedback to government institutions whose programs, especially in job training, are crucial to regional platforms.

For example, these networks might reveal the merits of one proposal for expanding and renaming the Manufacturing Extension Partnership (MEP) as the Innovation and Productivity Extension Partnership (IPEP). The MEP program is designed to enhance process innovation in SME manufacturers.

Rather than decide how to adjust, expand, and broaden this program top-down from Washington, such ideas could emerge from regional networks. As Breznitz and Murphree argue concerning China, regional experimentation produced ideas for reform and business that were later adopted nationwide.¹⁷³ The same principle of allowing local platforms to take the lead and then cautiously applying the models more broadly should be considered.

A third prong of “networked solution” institutions crosses over with creation of platform capabilities. Solving many production process problems through collaboration may lead to collective investments, perhaps through co-op systems, in certain kinds of capabilities. For example, most SMEs have limited capacity for original applications of ICT customized to their needs or design innovations. Whereas many firms offer to provide these inputs as outsourced activities, specialist suppliers often draw from a relatively small pool of relevant experiences. Providing a node at the regional level for comparing ICT and design ideas, and even generating new ones relevant to the cluster, could be powerful.

We strongly encourage the federal and state governments to open and quickly expand programs such as the traditional industries program of the Israeli chief scientist. Of particular relevance is the part of that program which matches graduate students from high-tech disciplines with production SMEs. Special attention should be given to how to incentivize actors (students and manager/owners of companies) so that these internships in

¹⁷²G. McDermott, R. A. Corredoira, and G. Kruse, “Public-Private Institutions as Catalysts of Upgrading in Emerging Markets,” *Academy of Management Journal* 52, no. 6 (2009): 1292.

¹⁷³B. Breznitz and M. Murphree, *Run of the Red Queen: Government, Innovation, Globalization and Economic Growth in China* (New Haven, CT: Yale University Press, 2011).

companies would lead to both innovative projects and outcomes and, at least as important, routinizing innovation activities in SMEs.

Our fourth implementing action is to align the incentives of public officials using a new set of metrics to judge success in building networks. Public officials respond strongly to incentive structures that for good or ill shape their behavior and thus their economic outcomes. In China, one reason local governments invest so heavily in new industrial parks and infrastructure construction is that these projects rapidly generate jobs, investment, and growth—three metrics on which promotion for officials is based.

For initiatives such as an “industrial extension service” or public-private dialogues and road mapping to work, we must find ways to motivate and evaluate public officials. Despite popular distrust of government in America, government officials at the regional level do valuable economic development work. Such officials have substantial levels of bipartisan support, but we must rethink many of the conventional measures used to gauge their success. Metrics such as counts of the number of firms created, or the number of new jobs created within a defined period of time, or even cost-benefit analysis, give exactly the wrong incentives to policy makers, and lead to wrong evaluation of the effectiveness of policies by politicians and the public.

As in China, such incentives would encourage short-term and direct activities rather than long-term and more complicated steps toward building a sustainable I&P innovation ecosystem. We must define metrics that actually measure the growth of networks, the diffusion of innovation within them, the percentage of new production technologies that are implanted in the United States, the growth rate of I&P innovation (currently not properly measured), and the growth of new high-end specialized producers in the United States.

V. Conclusion

With some prudent reinforcement, the United States has the attributes needed to continue to lead the world in novel-product innovation, an enormous plus for American prosperity. However, the weaknesses in the I&P innovation system must be fixed. Many of the biggest payoffs for employment will come from these repairs, and they will also reinforce US strengths in novel-product innovation in the long term. In tackling I&P innovation, there are real lessons that can be learned both from experimentation and experience abroad and from state and local level projects within the United States. These lessons need to be applied in order to encourage more production in the United States

and the enhancement and preservation of US capabilities in the critical, but underappreciated, field of incremental and process innovation.

This paper is built on the collective results of a Connect Innovation Institute (San Diego) project: "Innovation, Production, and Sustainable Job Creation: Reviving U.S. Prosperity." The project group consisted of Susan Helper, William Lazonick, Erica Fuchs, Josh Whitford, Elisabeth Reynolds, and Jenny Kuan, without whom this paper would not have been possible. We also want to thank Mark Muro, Howard Wail, and Bruce Katz of the Brookings Institution for comments and suggestions. This project would not have been possible without the leadership of Duane Roth and the active help and encouragement of Mary Walshock and Karen Winston at the Connect Innovation Institute. We are also particularly grateful to the generous funding of the Waitt Foundation, the Kauffman Foundation, and the National Science Foundation under grant SES-0964907 for their generous support of this research.

Supply Chains and Equitable Growth

Susan Helper

Frank Tracy Carlton Professor of Economics, Case Western University Weatherhead School of Management

The US economy has undergone a structural transformation in recent decades. Large firms have shifted from doing many activities in-house to buying goods and services from a complex web of other companies. These outside suppliers make components and provide services in areas such as logistics, cleaning, and information technology. Although this change in the structure of supply chains began decades ago, neither public policy nor business practice has adequately dealt with the challenges posed by this restructuring. As a result, weakness in supply chains threatens US competitiveness by undermining innovation and contributes to the erosion of US workers' standard of living. This essay suggests policies to promote supply chain structures that stimulate equitable growth—that is, policies that both promote innovation and also ensure that the gains from innovation are broadly shared.

The Role of Supply Chains in the US Economy

A supply chain links companies, often in multiple industries and multiple locations, to design, produce components, and assemble and distribute a final product, such as a car, a computer, or a restaurant meal.¹⁷⁴ For much of the 20th century, a significant part of the US economy was characterized by supply chains that were vertically integrated.¹⁷⁵ Beginning in the 1970s and 1980s, large firms in many industries began to sell off assets and outsource work. Today, a lead firm typically designs products and directs production by

Reprinted with the permission of the publisher, the Washington Center for Equitable Growth, 31 October 2016.

This article is part of our series: Delivering Equitable Growth: Strategies for the Next Administration.

¹⁷⁴Rashmi Banga, "Measuring Value in Global Value Chains" (working paper, Centre for WTO Studies, New Dehli, May 2013), <http://wtocentre.iift.ac.in/workingpaper/Measuring%20Value%20in%20Global%20Value%20Chains%20CWS%20WP%20Final.pdf>.

¹⁷⁵Alfred Chandler, *The Visible Hand* (Cambridge, MA: Harvard University Press, 1978).

multiple tiers of suppliers in many locations but does not own most of these suppliers.¹⁷⁶

Supply chains made up of these financially independent firms are now the largest driver of firms' costs. The average US-based multinational firm buys intermediate inputs that compose about 75% of the value of its output; a domestically owned firm buys intermediate inputs equal to about 50% of output value.¹⁷⁷ Contrary to the common impression, most of these suppliers are domestic, even in manufacturing.¹⁷⁸ These outsourced supply chains differ from vertical integration in that the lead firm does not own supplier facilities. The lead firm benefits from this arrangement by gaining access to products made by suppliers with experience in making similar products for multiple customers and by not being responsible for subsidiaries' fixed costs.

These supply chains also typically differ from economists' model of perfect competition, in which transactions between firms are at arm's length and the only information that crosses firm boundaries is price information. In contrast, many suppliers make products specifically tailored to meet the needs of the lead firm and frequently exchange information with the lead firm regarding designs, production processes, and future plans. Lead firms find this arrangement advantageous because they are able to quickly obtain components tailored to their specific needs. The complementary disadvantage is that firms are often unable to change suppliers easily.

On the one hand, sharing suppliers with other lead firms has significant benefits, such as shared knowledge across customers and reduced fixed costs. On the other hand, lead firms may lack incentive to invest in upgrading the supplier's capabilities if that supplier may also use those capabilities to serve a competitor. Firms' success depends on having robust networks of suppliers, but no one firm is responsible for keeping these networks healthy.

¹⁷⁶Marcel P. Timmer, Abdul Azeez Erumban, Bart Los, Robert Stehrer, and Gaaitzen J. de Vries, "Slicing Up Global Value Chains," *Journal of Economic Perspectives* 28, no. 2 (2014), <https://www.aeaweb.org/articles?id=10.1257/jep.28.2.99>. See also Richard Baldwin and Javier Lopez-Gonzalez, "Supply-Chain Trade: A Portrait of Global Patterns and Several Testable Hypotheses" (NBER Working Paper 18957, April 2013), <http://www.nber.org/papers/w18957>.

¹⁷⁷James Fetzer and Erich H. Strassner, "Identifying Heterogeneity in the Production Components of Globally Engaged Business Enterprises in the United States," Bureau of Economic Analysis (Washington, DC: US Department of Commerce, 2015), <https://bea.gov/about/pdf/acm/2015/november/identifying-heterogeneity-fetzer-strassner.pdf>.

¹⁷⁸Jessica R. Nicholson and Ryan Noonan, "What Is Made in America?" Economics and Statistics Administration (Washington, DC: US Department of Commerce, 2014), http://www.esa.doc.gov/sites/default/files/whatismadeinamerica_0.pdf.

Implications of Supply Chain Structure for Innovation

Because innovation is concentrated in manufacturing—two-thirds of private-sector research and development is performed in manufacturing—this section looks at supply chains in manufacturing only (data are not readily available for innovation in other sectors).¹⁷⁹

Firms with fewer than 500 employees are an increasing share of manufacturing employment, accounting for 42% of such workers in 2012. These small firms struggle at each phase of the innovation process. They are only 15% as likely to conduct research and development as large firms. Small firms also struggle to obtain financing and a first customer to help them commercialize a new product or process. Finally, small manufacturers have trouble adopting new products or processes developed by others, due to difficulty in learning about and financing new technology. As a result, small manufacturers are only 60% as productive as large firms.¹⁸⁰

A skeptic may ask why large lead firms cannot innovate enough to support their entire production network. But problems such as reducing the vibration of a wind turbine require holistic problem-solving; a machine composed of many parts that exert strong forces on each other cannot simply be divided into one problem for the gearbox manufacturer to solve, one for the rotor manufacturer to solve, and another for the assembly team to solve. Limiting innovation to lead firms deprives the supply chain of insights that come from being very close to a particular type of production or use.¹⁸¹ In addition, long-term supplier-customer relationships built on trust and collaboration best facilitate progress toward these goals; lack of such relationships accounts for many of the problems industries face in moving new technologies from lab to market.

Implications of Supply Chain Structure for Job Quality

Workers are employed in supply chains in a variety of ways. Instead of being hired directly by lead firms as regular employees, workers may be hired by temporary help agencies and are often referred to as “contingent workers.”

¹⁷⁹Executive Office of the President and the US Department of Commerce, “Supply Chain Innovation: Strengthening America’s Small Manufacturers,” March 2015, https://www.whitehouse.gov/sites/default/files/docs/supply_chain_innovation_report.pdf.

¹⁸⁰Executive Office of the President and the US Department of Commerce, “Supply Chain Innovation: Strengthening America’s Small Manufacturers.”

¹⁸¹Gary P. Pisano and Willy C. Shih, “Restoring American Competitiveness,” *Harvard Business Review*, 2009, <https://hbr.org/2009/07/restoring-american-competitiveness/ar/1>. See also Suzanne Berger with the MIT Task Force on Production in the Innovation Economy, *Making in America: From Innovation to Market* (Cambridge, MA: MIT Press, 2013).

Alternatively, they may be hired as regular workers at supplier firms or as independent contractors.

A variety of studies find that these forms of outsourcing of employment, especially as carried out in the United States, typically create undesirable outcomes for workers in areas such as wages, benefits, job security, and safety.¹⁸² Contingent workers earn 10.5% less per hour and 47.9% less per year than non-contingent workers, and are more likely to suffer workplace injury.¹⁸³ Workers employed at suppliers, even as regular workers, generally earn less than workers at lead firms, which tend to be larger.

Wages are typically lower at suppliers than at lead firms because of the barriers to innovation discussed above, which reduce productivity; the absence of pressures to reduce wage differentials within a firm due to norms of fairness; and greater pressure on wages at outside suppliers, which are more easily replaced than are internal divisions.

Market and Network Failures in Supply Chains

Three forms of market failure contribute to the central tendency of US supply chains to suppress innovation and make jobs worse:

- *Free-rider problems between firms.* When a lead firm makes investments in upgrading its suppliers—by providing technical assistance to suppliers, training supplier workers, or helping them invest in new equipment—some of this improved capability will often spill over to benefit a supplier’s other customers, including the lead firm’s rivals. Lead firms thus have less incentive to invest in their suppliers than would be socially beneficial.¹⁸⁴
- *Silos within firms.* Internal conflicts between departments within a lead firm can mean a focus on finding suppliers with low prices rather than on those providing high quality and innovation. An easy way for firms to evaluate their purchasing departments, for example, is the extent to

¹⁸²Annette Bernhardt, Rosemary Batt, Susan Houseman, and Eileen Appelbaum, “Domestic Outsourcing in the U.S.: A Research Agenda to Assess Trends and Effects on Job Quality” (IRLE Working Paper 102-16, February 2016), <http://irle.berkeley.edu/workingpapers/102-16.pdf>.

¹⁸³US Government Accountability Office, *Contingent Workforce: Size, Characteristics, Earnings, and Benefits*, GAO-15-168R (Washington, DC: Government Accountability Office, 20 April 2015), <http://www.gao.gov/assets/670/669766.pdf>.

¹⁸⁴Josh Whitford, *The New Old Economy: Networks, Institutions, and the Organizational Transformation of American Manufacturing* (Oxford, UK: Oxford University Press, 2006); Susan Helper and Janet Kiehl, “Developing Supplier Capabilities: Market and Non-Market Approaches,” *Industry and Innovation* 11, no. 1–2 (2004): 89–107, <http://faculty.weatherhead.case.edu/helper/papers/helperkiehl.pdf>.

which they reduce the price per unit they buy. A purchasing agent could thus be rewarded for choosing a supplier whose costs are \$1,000 less than a rival supplier's—even if that supplier's skimping on quality control later causes the shutdown of a production line that costs the operations department \$100,000. It may seem unlikely that sophisticated companies would fall prey to such problems, but quality and innovation are harder to measure than prices, and their benefits often accrue to departments other than purchasing.¹⁸⁵

- *Profit protection.* Outsourcing of work often reduces workers' access to profits earned by the lead firm. Organizational structures tend to minimize wage differentials within firms, due to both norms of fairness and to a desire to promote cooperation within an organization. Firms with a high degree of market power have lots of profits to protect, which they often do by adopting policies that make their suppliers interchangeable, even at a cost to efficiency.¹⁸⁶

The result of these market failures is an emphasis in the United States on arm's length rather than collaborative governance of supply chains, and a hollowing out of productive ecosystems, as firms set up incentives for their purchasing departments that privilege supplier firms that can win competitive bidding wars. These "winners" tend to be small firms with low expenditures on overhead costs, covering such things as salaries for managers and engineers and worker training. In extreme cases, such as garment production or janitorial services, competition is so fierce that firms compete in part by violating laws on safety, minimum wages, overtime, and disposal of toxic waste. In the rare instances in which these firms are caught, they often can file for bankruptcy and reopen under another name.¹⁸⁷

¹⁸⁵Whitford, *The New Old Economy: Networks, Institutions, and the Organizational Transformation of American Manufacturing*, 2006; Susan Helper and Rebecca Henderson, "Management Practices, Relational Contracts, and the Decline of General Motors," *Journal of Economic Perspectives* 28, no. 1 (2014): 49–72, <https://www.aeaweb.org/articles?id=10.1257/jep.28.1.49>.

¹⁸⁶A medium amount of market power can, however, promote efficient collaboration in supply chains; if the lead firm has no economic profits, it may be unable to make commitments that promote long-term, mutually profitable relationships. Susan Helper and David I. Levine, "Long-Term Supplier Relations and Product-Market Structure," *Journal of Law, Economics & Organization* 8, no. 3 (1992): 561–81, https://www.researchgate.net/profile/David_Levine6/publication/5214362_Long-Term_Supplier_Relations_and_Product-Market_Structure/links/0912f5106eeeba303600000.pdf.

¹⁸⁷David Weil, *The Fissured Workplace* (Cambridge, MA: Harvard University Press, 2014).

Policies to Promote Innovative Supply Chains with Good Jobs

Outsourcing has its advantages, principally in making possible a potentially efficient division of labor in which specialist firms can achieve economies of scale and diffuse best practices by serving a variety of customers. Yet lead firms' zealous embrace of the non-collaborative version of this strategy has resulted in significant weaknesses in innovation and job quality in the United States.

Tackling these challenges will help address some root causes of wage inequality and productivity stagnation in US manufacturing and service industries. Policies in five areas will help:

Encourage Firms to Adopt Collaborative Supply Chain Practices.

Public support for economic growth has long focused on the diffusion of physical technologies, yet the diffusion of operational insights may be just as valuable. Evidence suggests supply chains with more collaborative practices are more innovative.¹⁸⁸ The next administration should use its convening power to encourage lead firms to take steps such as these:

- *Offer suppliers' assurance that they will receive a fair return on investments they make in new technologies and in upgrading their capabilities.* In order to become partners in innovation, suppliers need to develop better capabilities in product and process design and to upgrade equipment.
- *Promote information-sharing and make changes in their own operations as a result of supplier suggestions.* A key insight from the Toyota production system is that firms and workers who are close to production have access to information not easily available to those at the top of the chain.¹⁸⁹ Firms that establish mechanisms to learn from their suppliers can significantly improve cost and quality.
- *Use a "total cost of ownership" approach when making purchasing decisions.* Firms should consider impacts of sourcing decisions on quality and

¹⁸⁸Executive Office of the President and the US Department of Commerce, "Supply Chain Innovation: Strengthening America's Small Manufacturers."

¹⁸⁹In contrast, Taylorist theories of management hold that "brain work" (such as process design) and "hand work" (such as production) should be done by different people, with little payoff to feedback between the groups. Susan Helper and Rebecca Henderson, "Management Practices, Relational Contracts, and the Decline of General Motors," *Journal of Economic Perspectives* 28, no. 1 (2014): 49–72, <https://www.aeaweb.org/articles?id=10.1257/jep.28.1.49>; Paul S. Adler and Bryan Borys, "Two Types of Bureaucracy: Enabling and Coercive," *Administrative Science Quarterly* 41, no. 1 (1996): 61–89, <https://www.jstor.org/stable/2393986>.

innovation as well as on price per unit purchased.¹⁹⁰ Forming long-term, collaborative relationships with highly competent suppliers may be in a firm's best overall interest, yet purchasing departments are not always incentivized to consider these benefits.

Nurture Productive Ecosystems of Firms, Universities, Communities, and Unions. One reason for the struggles that small- and medium-sized US firms face is that they are “home alone,” with few institutions to help with innovation, training, and finance.¹⁹¹ For reasons of both equity and efficiency, these firms should not depend solely on their customers for strategic support.

Policies that nurture small firms, local universities, their communities, and unions could help the firms leverage their advantages over their larger brethren in nimbleness and strong community ties. Germany's Mittelstand (medium-sized firms) are the backbone of the German manufacturing sector due to the help they get from community banks, applied research institutes, and unions.¹⁹² In the United States, the unionized construction sector has developed structures that create good jobs and fast diffusion of new techniques, even though the industry remains characterized by small firms and work that is often intermittent. Building trades unions work with signatory employers to provide apprenticeships, continuing education programs, and portable benefits.¹⁹³

Federal technology assets should be better deployed as well, continuing the work begun by the Obama White House Supply Chain Innovation

¹⁹⁰US Department of Commerce, “Assess Costs Everywhere” (tool for assessing sourcing costs), at <http://acetool.com-merce.gov>.

¹⁹¹Suzanne Berger with the MIT Task Force on Production in the Innovation Economy, *Making in America: From Innovation to Market* (Boston: MIT Press, 2013). See also Robert D. Ezell and Stephen J. Atkinson, “International Benchmarking of Countries’ Policies and Programs Supporting SME Manufacturers” (Information Technology & Innovation Foundation, September 2011), <http://www.itif.org/files/2011-sme-manufacturing-tech-programs.pdf>; Susan Helper, Timothy Krueger, and Howard Wial, “Why Does Manufacturing Matter? Which Manufacturing Matters?” (Washington, DC: Brookings Institution, 2012), <http://www.brookings.edu/research/papers/2012/02/22-manufacturing-helper-krueger-wial>.

¹⁹²Susan Helper, Timothy Krueger, and Howard Wial, “Why Does Manufacturing Matter? Which Manufacturing Matters?” (Washington, DC: Brookings Institution, 2012), <http://www.brookings.edu/research/papers/2012/02/22-manufacturing-helper-krueger-wial>; Suzanne Berger, *How We Compete: What Companies around the World Are Doing to Make It in Today's Global Economy* (New York: Currency, 2005).

¹⁹³Cihan Bilginsoy, “The Hazards of Training: Attrition and Retention in Construction Industry Apprenticeship Programs,” *Industrial & Labor Relations Review* 57, no. 1 (2003): 54–67, <http://tcimass.org/sites/builtbest.prometheuslabor.com/files/Attrition%20and%20retention.pdf>.

Initiative.¹⁹⁴ National labs can be encouraged to work with small as well as large firms, for example, and the Manufacturing Extension Partnership can expand its efforts to work with entire supply chains (rather than firms one by one) to identify sources of inefficiency. A century ago, the federal government played this role in agriculture by funding land grant universities, which not only led to the creation of knowledge but also created durable networks of researchers and practitioners through which such knowledge could quickly spread.¹⁹⁵

Promote Formation of Supply Chains in Industries That Advance National Goals. The free-rider problems discussed above are likely to be particularly acute in forming collaborative supply chains for new products, such as improved solar panels or wind turbines. These industries face additional market failures leading to underinvestment in addressing climate change. The Obama administration's Clean Energy Manufacturing Initiative helps to move new technologies out of the laboratory and into production. It would be useful to explicitly address the incentive and information issues in supply chains for producing and installing these products. The next administration could convene firms throughout the supply chain to engage in value analysis to improve product designs, to uncover hidden pockets of inventory, and to adopt total-cost-of-ownership techniques.

Promote Good-Jobs and High-Road Strategies. Much research documents the ways that firms can utilize "high-road" policies or "good-jobs" strategies to tap the knowledge of all their workers to create innovative products and processes.¹⁹⁶ High-road firms remain in business while paying higher wages than their competitors because their highly skilled workers help these firms achieve high rates of innovation, quality, and fast response to unexpected situations. The resulting high productivity allows these firms to pay

¹⁹⁴The White House, "FACT SHEET: Convening Manufacturing Leaders to Strengthen the Innovative Capabilities of the U.S. Supply Chain, including Small Manufacturers" (fact sheet, July 9, 2015), <https://www.whitehouse.gov/the-press-office/2015/07/09/fact-sheet-convening-manufacturing-leaders-strengthen-innovative>.

¹⁹⁵Irwin Feller, Patrick Madden, Lynne Kalreider, Dan Moore, and Laura Sims, "The New Agricultural Research and Technology Transfer Policy Agenda," *Research Policy* 16 no. 6 (1987): 315–325, <http://www.sciencedirect.com/science/article/pii/0048733387900175>.

¹⁹⁶See Eileen Appelbaum, Jody Gittell, and Carrie Leana, "High-Performance Work Practices and Sustainable Economic Growth" (Employment Policy Research Network, March 2011), <http://50.87.169.168/OJS/ojs-2.4.4-1/index.php/EPRN/article/view/1890/1888> for a summary; and Justin Wolfers and Jan Zilinsky, "Higher Wages for Low-Income Workers Lead to Higher Productivity" (Peterson Institute for International Economics, January 13, 2015), <https://piie.com/blogs/realtime-economic-issues-watch/higher-wages-low-income-workers-lead-higher-productivity>.

high wages while still making profits that are acceptable to the firms' owners. Collaborative supply chain governance plays an important role in providing the stability needed to support these strategies, from which lead firms also benefit.

Disincentivize Low-Road Production Strategies. Even in collaborative scenarios, wages are often less than in the old vertically integrated model. The corrosion of labor union power enables outsourcing, and the increase in outsourcing has, in turn, further decreased workers' bargaining power.

Thus, as important as it is to "pave the high road," it is also important to "block the low road."¹⁹⁷ The Department of Labor has begun to take advantage of modern supply chains' emphasis on "just-in-time" delivery, recognizing that reduced inventories make regulators' threat to shut down suppliers for violation of wage and hour laws a more potent threat.¹⁹⁸ New policies could combine such sticks with some carrots. The federal government could offer technical assistance, for example, to help small garment manufacturers move away from the existing low-road model, in which ill-trained workers typically do one simple operation to a garment and then pass it on to the next worker.

Instead, these firms could adopt a more agile production recipe, one that involves more broadly trained and higher-paid workers collaborating in teams—a high-road model sustained by greater productivity and reduced lead times.

Government should implement collaborative supply chain practices within its own purchasing, building on the Obama administration's nascent efforts to measure total cost of ownership and to ban supply chains with recent violations of labor and other laws from selling to the government.¹⁹⁹

Current outsourcing practices allow lead firms and their suppliers to reap the benefit of paying workers only when needed, while the risks of being left without earnings are borne by workers. Several proposals could improve the balance here: encouraging work-sharing in downturns (which would make hiring regular workers less costly), continuing to improve the portability of benefits across firms, and promoting schedule stability.

¹⁹⁷Dan Luria and Joel Rogers, *Metro Futures: Economic Solutions for Cities and Their Suburbs* (Boston: Beacon Press, 1999).

¹⁹⁸David Weil, *The Fissured Workplace* (Cambridge, MA: Harvard University Press, 2014).

¹⁹⁹The White House, "FACT SHEET: Fair Pay and Safe Workplaces Executive Order" (Fact sheet, July 31, 2014), <https://www.white-house.gov/the-press-office/2014/07/31/fact-sheet-fair-pay-and-safe-workplaces-executive-order>.

Retooling Supply Chains for Equitable Growth

Decisions about how to structure supply chains matter greatly for working Americans, yet this topic rarely takes a front seat in policy discussions of how to address rising inequality and stagnating productivity. In order to promote equitable growth, policymakers must understand how the economic pie is created—not just how it is divided.

Fundamental changes in the way supply chains operate threaten US economic competitiveness by undermining innovation and erode American workers' economic security. The rise over the past few decades of supply chains with small, weak firms leads to an increased presence of firms that innovate less and pay less. It is unlikely and undesirable, however, that the United States would return to the often bureaucratic and stifling vertically integrated supply chains of the mid-20th century.

We can do better. This essay has outlined government and corporate policies to promote both more innovation and better job quality in supply chains. In particular, more collaborative supply chains and better-supported local ecosystems could significantly improve the viability of “good jobs” strategies. The way the economic pie is created affects the way it is divided.

(For more detail on these proposals and the analysis behind them, see Susan Helper and Timothy Krueger, “Supply Chains and Equitable Growth,” Washington Center for Equitable Growth, September 29, 2016.)

Last Word

The Economic Sky Is Not Falling

Deirdre McCloskey

Distinguished Professor of Economics, History, English, and Communication, the University of Illinois at Chicago

For reasons I don't understand, people simply love to be told that the sky is falling. Yet it seldom does. Actually, never.

For example, a gaggle of economists, such as Lawrence Summers, Erik Brynjolfsson, Andrew McAfee, Edmund Phelps, Jeffrey Sachs, Laurence Kotlikoff, Tyler Cowen, and Robert Gordon have argued recently that Europe and the United States, on the frontier of betterment, are facing a slowdown of new ideas, with a skill shortage. Technological unemployment and “uncompetitiveness” and sadly slow growth, it is said, will be the result.

Maybe. Yet in the past couple of centuries numerous other learned economists have predicted similar slowdowns, none of which happened. The Keynesian economists in the late 1930s and the 1940s were confident in their prediction along the same lines as the current pessimists of a world “stagflationism.” The prediction was instantly falsified by the continuing Great Enrichment.

Similarly, in the first three-quarters of the nineteenth century the classical economists, Marx among them, expected landlords, or in Marx's case capitalists, to engorge the national product. On Malthusian grounds they expected people to stay at the \$2 or \$3 a day in 2016 prices typical of human life since the caves. It didn't happen. British real, inflation-corrected income per head per day is now thirty times higher, and if allowing for improvements in quality of goods and services, it is close to one hundred times higher.

Contrary to recent alarms, even in the already rich countries the real income for the poor continues to grow, if corrected to allow for radically better goods and services. Thirty years ago hip-joint replacement was experimental. Now it's routine. Tires and autos were unreliable. Now they never wear out. Once nothing could be done about clinical depression. Now something can. Further, in terms of real comforts—a roof, heating, ample clothing, decent food, adequate education, effective medicine, long life—the income is more and more *equally* spread. *Pace* Piketty. Worldwide the poorest among us are getting richer.

Reprinted with edits from *Prospect Magazine*, March 2016, with the permission of the publisher, Oxford University Press.

The Italian economists Patrizio Pagano and Massimo Sbracia argue that failures of previous stagnations—proposed after every major recession, they note—failed not so much in the impossible task of anticipating wholly new technology as in not grasping the further rewards of existing technology, such as nowadays computers.

Joel Mokyr, a deep student of the history of technology, recently offered some persuasive assurances on the matter of slowdown, directed specifically at the sky-is-falling convictions of his colleague at Northwestern University, the gloomy Gordon. Mokyr argues that the sciences behind biology and computers and the study of materials promise gigantic enrichment.

As the historian, politician, and nineteenth-century liberal Thomas Babington Macaulay asked in 1830, “On what principle is it that, when we see nothing but betterment behind us, we are to expect nothing but deterioration before us?” He continued:

If we were to prophesy that in the year 1930 a population of fifty million, better fed, clad, and lodged than the English of our time, will cover these islands, that Sussex and Huntingdonshire will be wealthier than the wealthiest parts of the West Riding of Yorkshire now are, that machines constructed on principles yet undiscovered will be in every house, many people would think us insane.

Whiggish and bourgeois and progress-minded and vulgarly pro-betterment though Macaulay was, he was in his prediction exactly right, even as to the U.K. population in 1930. If one includes in 1930 the recently separated Republic of Ireland, he was off by less than 2%.

And even the pessimistic, anti-Whiggish economists such as Gordon—“gloomsters,” the headline writers call them—would not deny that we have before us fifty or a hundred years in which now middling and poor countries such as South Africa and Brazil and Haiti and Bangladesh will catch up to what is already, in the rich countries, a stunningly successful level of average real income.

The Nobelist Edmund Phelps, among the pessimists, believes that many rich countries lack dynamism. Some of Gordon’s proposed “headwinds” are of that character. It hasn’t happened yet, but let’s suppose the sky does fall on Europe and its offshoots. China and India, making up about four in ten of world population, are growing with notable dynamism at upward of 7 to 12% per person per year.

To appreciate what will happen over the next fifty or a hundred years if such growth continues, as there is every reason to think it will, it’s a good idea to learn the “Rule of 72.” The rule is that something (such as income) growing at 1% per year takes seventy-two years to double. (Rest assured, the fact is

not obvious without calculation. It just happens to be true. You can confirm it by taking out your calculator and multiplying 1.01 by itself seventy-two times.) It follows that if the something grows twice as fast, at 2% instead of 1%, the something will double, of course, in *half* the time, thirty-six years. A runner going twice as fast will arrive at the kilometer mark in half the time. Similarly, something growing at 3% a year will double in a third of the time, or twenty-four years.

We have in fact seen 1990-2016, even with the Great Recession, a real (constant 2011 dollars at purchasing power parity) growth rate worldwide of 2% per year per person, which is for example the average rate over two centuries in the USA.²⁰⁰ It will result in a doubling of the material welfare of the world's average person within a long generation ($72/2 = 36$ years), or two shorts, with economies of scale in world invention kicking up the rate. In two such generations, 72 years, it would mean a quadrupling, which would raise the average real income in the world at the end of this century to the levels attained in 2016 in the United States, a country that for well over a century has sustained the world's highest per-person income of any place larger than Norway. Pretty good. And it will be pretty good for solving many if not all of the problems in the soul and in the society and in the environment.

All the economists who have looked into the evidence agree that the average real income per person in the world is rising fast, and with every prospect of continuing tomorrow, and for the coming century or more. The result will be a gigantic increase in the number of scientists, designers, writers, musicians, engineers, entrepreneurs, and ordinary businesspeople devising betterments which will spill over to the now rich countries allegedly lacking in dynamism, or facing headwinds. Unless one believes in mercantilist/business-school fashion that a country must "compete" to prosper from world betterment, then, even the leaky boats of the Phelpsonian/Gordonesque "undynamic countries" will rise.

In short, no limit to fast world or U.S. or European growth of per-person income is close at hand, no threat to "jobs," no cause for pessimism—not in your lifetime, or that of your great-grandchildren. Then, in the year 2100, with everyone on the planet enormously rich by historical standards, and hundreds of times more scientists and entrepreneurs working on improvements in solar power and methane burning, we can reconsider the limits to growth, and the falling sky.

²⁰⁰<https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD>.

**The CFA Institute
Research Foundation
Board of Trustees
2019–2020**

Chair
Ted Aronson, CFA
AJO

Heather Brillant, CFA
Diamond Hill

Margaret Franklin, CFA
CFA Institute

Bill Fung, PhD
Aventura, FL

Daniel Gamba, CFA
BlackRock

JT Grier, CFA*
Virginia Retirement
System

Joanne Hill
CBOE Vest Financial

Roger Ibbotson*
Yale School of Management

Joachim Klement, CFA
Independent

Vikram Kuriyan, PhD, CFA
GWA and Indian School
of Business

Aaron Low, CFA
LUMIQ

Mauro Miranda, CFA
Panda Investimentos
AAI Ltda.

Lotta Moberg, CFA
William Blair

Sophie Palmer, CFA
Jarislowsky Fraser

Dave Uduanu, CFA
Sigma Pensions Ltd

*Emeritus

Officers and Directors

Executive Director
Bud Haslett, CFA
CFA Institute

Gary P. Brinson Director of Research
Laurence B. Siegel
Blue Moon Communications

Associate Research Director
Luis Garcia-Feijóo, CFA, CIPM
Coral Gables, Florida

Secretary
Jessica Lawson
CFA Institute

Treasurer
Kim Maynard
CFA Institute

Research Foundation Review Board

William J. Bernstein
Efficient Frontier
Advisors

Elroy Dimson
London Business School

Stephen Figlewski
New York University

William N. Goetzmann
Yale School of
Management

Elizabeth R. Hilpman
Barlow Partners, Inc.

Paul D. Kaplan, CFA
Morningstar, Inc.

Robert E. Kiernan III
Advanced Portfolio
Management

Andrew W. Lo
Massachusetts Institute
of Technology

Alan Marcus
Boston College

Paul O'Connell
FDO Partners

Krishna Ramaswamy
University of
Pennsylvania

Andrew Rudd
Advisor Software, Inc.

Stephen Sexauer
Allianz Global Investors
Solutions

Lee R. Thomas
Pacific Investment
Management Company

Named Endowments

The CFA Institute Research Foundation acknowledges with sincere gratitude the generous contributions of the Named Endowment participants listed below.

Gifts of at least US\$100,000 qualify donors for membership in the Named Endowment category, which recognizes in perpetuity the commitment toward unbiased, practitioner-oriented, relevant research that these firms and individuals have expressed through their generous support of the CFA Institute Research Foundation.

Ameritech	Miller Anderson & Sherrerd, LLP
Anonymous	Nikko Securities Co., Ltd.
Robert D. Arnott	Nippon Life Insurance Company of Japan
Theodore R. Aronson, CFA	Nomura Securities Co., Ltd.
Asahi Mutual Life Insurance Company	Payden & Rygel
Batterymarch Financial Management	Provident National Bank
Boston Company	Frank K. Reilly, CFA
Boston Partners Asset Management, L.P.	Salomon Brothers
Gary P. Brinson, CFA	Sassoon Holdings Pte. Ltd.
Brinson Partners, Inc.	Scudder Stevens & Clark
Capital Group International, Inc.	Security Analysts Association of Japan
Concord Capital Management	Shaw Data Securities, Inc.
Dai-Ichi Life Insurance Company	Sit Investment Associates, Inc.
Daiwa Securities	Standish, Ayer & Wood, Inc.
Mr. and Mrs. Jeffrey Diermeier	State Farm Insurance Company
Gifford Fong Associates	Sumitomo Life America, Inc.
Investment Counsel Association of America, Inc.	T. Rowe Price Associates, Inc.
Jacobs Levy Equity Management	Templeton Investment Counsel Inc.
John A. Gunn, CFA	Frank Trainer, CFA
John B. Neff, CFA	Travelers Insurance Co.
Jon L. Hagler Foundation	USF&G Companies
Long-Term Credit Bank of Japan, Ltd.	Yamaichi Securities Co., Ltd.
Lynch, Jones & Ryan, LLC	
Meiji Mutual Life Insurance Company	

Senior Research Fellows

Financial Services Analyst Association

For more on upcoming Research Foundation publications and webcasts, please visit www.cfainstitute.org/learning/foundation.



CFA Institute
Research
Foundation

Available online at www.cfainstitute.org

ISBN 978-1-944960-83-4



9 781944 960834 >