Federal Reserve Bank of St. Louis

REVIEW

JULY/AUGUST 2009 VOLUME 91, NUMBER 4

Projecting Potential Growth: Issues and Measurement

Proceedings of the Thirty-Third Annual Economic Policy Conference of the Federal Reserve Bank of St. Louis

What Do We Know (And Not Know) About Potential Output?
Susanto Basu and John G. Fernald

Issues on Potential Growth Measurement and Comparison: How *Structural* Is the Production Function Approach?
Christophe Cahn and Arthur Saint-Guilhem

Parsing Shocks: Real-Time Revisions to Gap and Growth Projections for Canada
Russell Barnett, Sharon Kozicki, and Christopher Petrinec

The Challenges of Estimating Potential Output in Real Time
Robert W. Arnold

Trends in the Aggregate Labor Force
Kenneth J. Matheny

Potential Output in a Rapidly Developing Economy: The Case of China and a Comparison with the United States and the European Union
Jinghai Zheng, Angang Hu, and Arne Bigsten

Estimating U.S. Output Growth with Vintage Data in a State-Space Framework
Richard G. Anderson and Charles S. Gascon

Panel Discussion
Carlos Hamilton Araujo, Seppo Honkapohja, and James Bullard

Commentaries
Rodolfo E. Manuelli, Jon Faust, Gregor W. Smith, Robert J. Tetlow, Ellis W. Tallman, Xiaodong Zhu, Dean Croushore

REVIEW

Director of Research
Christopher J. Waller

Senior Policy Adviser
Robert H. Rasche

Deputy Director of Research
Cletus C. Coughlin

Review *Editor-in-Chief*
William T. Gavin

Research Economists
Richard G. Anderson
Alejandro Badel
Subhayu Bandyopadhyay
Silvio Contessi
Riccardo DiCecio
Thomas A. Garrett
Carlos Garriga
Massimo Guidolin
Rubén Hernández-Murillo
Luciana Juvenal
Kevin L. Kliesen
Natalia A. Kolesnikova
Michael W. McCracken
Christopher J. Neely
Michael T. Owyang
Michael R. Pakko
Adrian Peralta-Alva
Rajdeep Sengupta
Daniel L. Thornton
Howard J. Wall
Yi Wen
David C. Wheelock

Managing Editor
George E. Fortier

Editors
Judith A. Ahlers
Lydia H. Johnson

Graphic Designer
Donna M. Stiller

The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Projecting Potential Growth: Issues and Measurement

President's Welcome 179
James Bullard

Editor's Introduction 181
Richard G. Anderson

187
What Do We Know (And Not Know) About Potential Output?
Susanto Basu and John G. Fernald

Commentary 215
Rodolfo E. Manuelli

221
Issues on Potential Growth Measurement and Comparison: How *Structural* Is the Production Function Approach?
Christophe Cahn and Arthur Saint-Guilhem

Commentary 241
Jon Faust

247
Parsing Shocks: Real-Time Revisions to Gap and Growth Projections for Canada
Russell Barnett, Sharon Kozicki, and Christopher Petrinec

Commentary 267
Gregor W. Smith

271
The Challenges of Estimating Potential Output in Real Time
Robert W. Arnold

Commentary 291
Robert J. Tetlow

297
Trends in the Aggregate Labor Force
Kenneth J. Matheny

Commentary 311
Ellis W. Tallman

317
Potential Output in a Rapidly Developing Economy: The Case of China and a Comparison with the United States and the European Union
Jinghai Zheng, Angang Hu, and Arne Bigsten

Commentary 343
Xiaodong Zhu

349
Estimating U.S. Output Growth with Vintage Data in a State-Space Framework
Richard G. Anderson and Charles S. Gascon

Commentary 371
Dean Croushore

Panel Discussion 383
Carlos Hamilton Araujo, Seppo Honkapohja, and James Bullard

Review is published six times per year by the Research Division of the Federal Reserve Bank of St. Louis and may be accessed through our website: **research.stlouisfed.org/publications/review**. All nonproprietary and nonconfidential data and programs for the articles written by Federal Reserve Bank of St. Louis staff and published in *Review* also are available to our readers on this website. These data and programs are also available through Inter-university Consortium for Political and Social Research (ICPSR) via their FTP site: www.icpsr.umich.edu/pra/index.html. Or contact the ICPSR at P.O. Box 1248, Ann Arbor, MI 48106-1248; 734-647-5000; **netmail@icpsr.umich.edu**.

Single-copy subscriptions are available free of charge. Send requests to: Federal Reserve Bank of St. Louis, Public Affairs Department, P.O. Box 442, St. Louis, MO 63166-0442, or call (314) 444-8809.

General data can be obtained through FRED (Federal Reserve Economic Data), a database providing U.S. economic and financial data and regional data for the Eighth Federal Reserve District. You may access FRED through our website: **research.stlouisfed.org/fred**.

Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Please send a copy of any reprinted, published, or displayed materials to George Fortier, Research Division, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166-0442; **george.e.fortier@stls.frb.org**. Please note: Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis. Please contact the Research Division at the above address to request permission.

© 2009, Federal Reserve Bank of St. Louis.

ISSN 0014-9187

Contributing Authors

Richard G. Anderson
Federal Reserve Bank of St. Louis
richard.g.anderson@stls.frb.org

Carlos Hamilton Araujo
Central Bank of Brazil
carlos.araujo@bcb.gov.br

Robert W. Arnold
Congressional Budget Office
bob.arnold@cbo.gov

Russell Barnett
Bank of Canada
rbarnett@bankofcanada.ca

Susanto Basu
Boston College
susanto.basu@bc.edu

Arne Bigsten
Gothenburg University
arne.bigsten@economics.gu.se

James Bullard
Federal Reserve Bank of St. Louis
james.b.bullard@stls.frb.org

Christophe Cahn
Paris School of Economics
christophe.cahn@pse.ens.fr

Dean Croushore
University of Richmond
dcrousho@richmond.edu

Jon Faust
Johns Hopkins University
faustj@jhu.edu

John G. Fernald
Federal Reserve Bank of San Francisco
john.fernald@sf.frb.org

Charles S. Gascon
Federal Reserve Bank of St. Louis
charles.s.gascon@stls.frb.org

Seppo Honkapohja
Board of the Bank of Finland
seppo.honkapohja@bof.fi

Angang Hu
Tsinghua University
anganghu@mail.tsinghua.edu.cn

Sharon Kozicki
Bank of Canada
skozicki@bankofcanada.ca

Rodolfo E. Manuelli
Washington University in St. Louis
manuelli@artsci.wustl.edu

Kenneth J. Matheny
Macroeconomic Advisers
matheny@macroadvisers.com

Christopher Petrinec
Bank of Canada
cpetrinec@bankofcanada.ca

Arthur Saint-Guilhem
European Central Bank
arthur.saint-guilhem@ecb.int

Gregor W. Smith
Queen's University
smithgw@econ.queensu.ca

Ellis W. Tallman
Oberlin College
etallman@oberlin.edu

Robert J. Tetlow
Federal Reserve Board
robert.j.tetlow@frb.gov

Jinghai Zheng
Gothenburg University
Jinghai.Zheng@economics.gu.se

Xiaodong Zhu
University of Toronto
xzhu@chass.utoronto.ca

President's Welcome

James Bullard

As president of the Bank, it is my pleasure to welcome you to the Thirty-Third Annual Policy Conference of the Federal Reserve Bank of St. Louis.

This conference concerns measurement of the economy's potential output. The concept of potential output is straightforward to define—the economy's maximum sustained level of output—but difficult to measure. Inclusion of the term *sustained* suggests that the concept of potential growth is closely tied to inflation—a low, stable inflation rate is essential if an economy is to attain maximum economic growth and, hence, remain through time at or near its potential level of output.

In macroeconomic stabilization theory and practice, the concept of potential growth has a long history. Early analyses focused on the *output gap*. Fortunately, belief in an exploitable long-run tradeoff between the unemployment rate and the rate of inflation was rejected by economists decades ago. Today's classical and New Keynesian models suggest that, given enough time for adjustment and a benign pattern of shocks, the economy will adjust in the long run toward its potential level of output. The speed of such adjustment depends on the relative flexibility or inflexibility of wages, prices, and expectations—aptly summarized by Keynes's quip that "In the long run, we are all dead." But, taken literally, Keynes's call to action, as we now recognize, can be quite dangerous when near-term preliminary data contain significant uncertainty and measurement error, as demonstrated by the papers of Athanasios Orphanides, John Williams, and Simon van Norden (e.g., Orphanides and van Norden, 2002; and Orphanides and Williams, 2005).

The concept of potential output is an important feature of monetary policymaking. At our conference in 2007 in honor of Bill Poole, Lars Svensson and Noah Williams (2008, p. 275) characterized the task of policymakers as seeking to "navigate the sea of uncertainty." Correct economic stabilization policy, like correct navigation, requires a focus on the destination, or long-run objective. The Federal Reserve, in particular, operates with a dual mandate from the Congress to achieve both price stability and maximum employment. These goals are not in conflict—both require fostering an environment to support maximum sustainable growth. Academic policy models, while differing one from another, typically include a concept of potential output. Fixed-parameter policy rules, such as the Taylor rule, feature an output gap. Flexible inflation targeting models, such as those of Lars Svensson (e.g., Svensson, 1997) emphasize that inflation can, and does, at times, move away from the desired level. Thus, the choice of an optimal policy that will return inflation to its target depends on a tradeoff between the costs of the higher-but-falling inflation and any induced output gap (i.e., an output gap judged relative to some measure of potential output). One lesson of such models is that, even when monetary policymakers focus solely on achieving price

James Bullard is president of the Federal Reserve Bank of St. Louis.

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 179-80.
© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the FOMC. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

stability, the path of the output gap will enter into their deliberations regarding an optimal policy to reach that goal.

It is in this spirit of the important policy role of potential output that I welcome the speakers who will share their thoughts with us. We are particularly rich in speakers from abroad, bringing a distinct international focus to our discussions. I trust we will all increase our understanding of the concept of potential output and its role in policymaking.

REFERENCES

Orphanides, Athanasios and van Norden, Simon. "The Unreliability of Output Gap Estimates in Real Time." *Review of Economics and Statistics*, November 2002, *84*(4), pp. 569-83.

Orphanides, A. and Williams, John C. "Expectations, Learning, and Monetary Policy." *Journal of Economic Dynamics and Control*, November 2005, *29*(11), pp. 1807-08.

Svensson, Lars E.O. "Optimal Inflation Targets, 'Conservative' Central Banks, and Linear Inflation Contracts." *American Economic Review*, March 1997, *87*(1), pp. 98-114.

Svensson, Lars E.O. and Williams, Noah. "Optimal Monetary Policy Under Uncertainty: A Markov Jump-Linear-Quadratic Approach." Federal Reserve Bank of St. Louis *Review*, July/August 2008, *90*(4), pp. 275-93.

Editor's Introduction

Richard G. Anderson

The Thirty-Third Annual Policy Conference of the Federal Reserve Bank of St. Louis was held October 20-21, 2008. Since their inception the Bank's annual conferences have been labeled *policy* conferences—yet, they also are *research* conferences. Each conference has focused on an aspect of contemporary economic research that is useful to policymakers.

The 2008 conference explored the concept and measurement of "potential" output. This concept is central to macroeconomic policymaking because policy decisions are driven, at least in part, by estimates of two concepts: "Is real output greater or less than potential?" and "Is real output growing more or less rapidly than potential?" The answers to these questions are central to policy decisions.

The output gap features prominently in the famous Taylor rule. It also is featured in inflation-forecasting policy models of the type popularized by Lars Svensson, both as an input to the inflation-forecasting process and as part of the policy authority's loss function.

WHAT DO WE KNOW (AND NOT KNOW) ABOUT POTENTIAL OUTPUT?

The first two conference papers address the concept of potential in the context of Real Business Cycle and New Keynesian macroeconomic models. The first paper, presented by Susanto Basu and John Fernald, addresses our knowledge in models of the concept of potential output. The authors focus on two concepts—(i) potential output as the long-run trend in output, largely ignoring fluctuations around the trend; and (ii) potential output as the level of output attainable today in two cases: first, if wages and price are fully flexible, that is, adjust immediately to their frictionless equilibrium values, and second, if wages and prices are sticky, that is, slow to adjust to the new equilibrium values. The authors argue that the first concept may be useful in studies of long-run economic growth and development, where fluctuations around the trend are of secondary interest, but it is inappropriate to *assume* such behavior in the study of business cycles. The second concept focuses on technology shocks to the economy, that is, shocks that affect productivity either within individual firms or in entire markets. Generally, a negative shock reduces the level of potential output because it reduces the maximum amount of capital that the market economy, in general equilibrium and with flexible prices, will *choose* to use in production. Necessarily in this case, if prices and wages are fully flexible, actual output will track potential—the economy is in a continuous general equilibrium position conditional on the amount of physical capital available—and hence the gross domestic product (GDP) gap fails to be a well-defined concept in such models. When price and wage stickiness exists in a New Keynesian–style model, a negative shock decreases

Richard G. Anderson is a vice president and economist at the Federal Reserve Bank of St. Louis.

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 181-85.
© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

actual output. If potential is taken to be the flexible-price, continuous-optimization level of output, then the GDP gap is positive. The authors conclude that a two-sector, rather than one-sector, model is necessary to capture the economy's behavior because technology trends and shocks have different effects on the production of consumer and producers' durable goods. In his commentary, Rodolfo Manuelli lauds the direction of the paper—toward making more explicit in the context of models those assumptions that underlie policy analysis—but argues the capital deepening suggested by the Basu and Fernald model is fragile: The model's correspondence to the data varies across time periods. Further, small changes in production technologies between sectors (indexed by the capital share of output) can easily worsen the model's correspondence to observed data. In general, he argues, Basu and Fernald's results will remain difficult to interpret until embedded in a more detailed model.

ISSUES ON POTENTIAL GROWTH MEASUREMENT AND COMPARISON: HOW *STRUCTURAL* IS THE PRODUCTION FUNCTION APPROACH?

The second of the conference's theoretical papers was presented by Christophe Cahn and Arthur Saint-Guilhem. In an ambitious analysis, they embed a production function definition of potential output within a large-scale dynamic stochastic general equilibrium (DSGE) model and calibrate separate versions of the model to U.S. and euro-area data. The authors seek to provide a quantitative and comparative assessment of two approaches to potential output measurement, specifically, the DSGE model and the traditional Solow-style production function approach. The authors conclude that production function approaches are likely to overstate the role of structural factors in explaining cross-country differences in potential growth, relative to the differences in shocks across countries through time. In his commentary, Jon Faust emphasizes the value to policymakers that will arise as newer macroeconomic techniques, including DSGE models, are integrated into (but do not replace) a policy-formulation process based on older, more traditional concepts. Further, DSGE models likely will not be brought into the policy process until economists can compare and contrast within the models both production function–based and Real Business Cycle flexible price concepts.

PARSING SHOCKS: REAL-TIME REVISIONS TO GAP AND GROWTH PROJECTIONS FOR CANADA

Two papers at the conference (this paper and Anderson and Gascon, below) used vintage "real-time" data in their measures of potential output. In the first of these papers, Russell Barnett, Sharon Kozicki, and Christopher Petrinec of the Bank of Canada explore a new database that contains both vintage data on observed (published) real output and projections by Bank of Canada staff for 1994-2005. The output gap has played an important role in the conduct of monetary policy discussions at the Bank of Canada, largely within a New Keynesian framework. Vintage data are crucial to analysis and policymaking in this framework because, as Basu and Fernald emphasize in their contribution, the behavior of inflation in such models is determined by next-period expected inflation and by the gap between real output and potential output—measures strongly affected by data revisions. The authors focus on the dynamic updating process in which shocks and revisions to observed real output cause changes to projections of both future output and the output gap. In his commentary, Gregor Smith emphasizes that the Bank of Canada's process through which potential output entered policy discussions had two distinct segments: the extended multivariate filter and the Quarterly Projection Model. An identification issue arises regarding when and to what degree estimates of potential should be smoothed: Should estimates of potential be smooth and the response of policymakers to output gaps rapid, or should estimates of potential be volatile and the response of policymakers muted?

THE CHALLENGES OF ESTIMATING POTENTIAL OUTPUT IN REAL TIME

Practical monetary policymaking requires empirical measures—and forecasts—of potential output. Two of the conference's papers addressed this essential issue. In the first of these, Robert Arnold of the Congressional Budget Office (CBO) explains its methods for measuring and projecting potential output. The CBO's published measure is the "gold standard" among policymakers and academic economists and likely is the most common measure used in empirical research. Arnold emphasizes that potential is not a technological ceiling on the economy's production; rather, it is a measure of the maximum *sustainable* output. Attempts by policymakers to push actual output above potential output will result in supply-side pressures on labor, natural resource markets, and capital, culminating in unacceptable inflation. In addition to the maximum long-run sustainable output level, Arnold argues that potential output simultaneously should be an estimate of the trend in GDP. In combining these two requirements, Arnold imparts a distinctly long-run flavor to his measure—a flavor that tastes of both the two potential output concepts discussed by Basu and Fernald and Cahn and Saint-Guilhem in this issue. The principal driving factors of potential output are labor force growth, capital investment (capital deepening), and gains in total factor productivity. In his commentary, Robert Tetlow notes that Arnold's analysis is similar to an analysis on the same topic published in 1979 in the *Carnegie-Rochester Conference Series* and asks why 30 years of econometric and macroeconomic research appear to have affected the CBO's procedures so little. His conjectures are two. First, the CBO's underlying macro framework is distinctly Keynesian; that is, the majority of economic fluctuations come from shocks to aggregate demand, not shocks to technology and aggregate supply. Second, a large and detailed framework built within the Keynesian paradigm assists the CBO in answering in short order a wide range of questions posed by policymakers.

TRENDS IN THE AGGREGATE LABOR FORCE

The second paper of the conference to address empirical measurement of U.S. potential output was presented by Ken Matheny, senior economist at the St. Louis forecasting and consulting firm Macroeconomic Advisers. Matheny explains why trend growth in the labor force is the key determinant of trends in employment and in potential GDP. Perhaps the most difficult aspect of such projections is tying labor force participation rates to demographics, including fertility and current life expectancy. Matheny explains why recent highly pessimistic forecasts might be too gloomy and why labor force growth might be stronger than anticipated. Among the important contributing factors are life expectancy, household net worth, and the unemployment rate. The model suggests a pronounced upward shift in the labor force after 2011, due largely to higher-than-anticipated participation rates among older persons. If this occurs, not only will potential output be higher than anticipated but pressures will be relieved on government programs such as Social Security and Medicare. Matheny suggests that the impact could be as large as an additional half of 1 percentage point on potential output growth—but, when all things are considered, through 2017 Matheny's new forecast for potential output growth is only two-tenths of 1 percentage point—2.6 percent per annum versus 2.4 percent—greater than the CBO forecast. In his commentary, Ellis Tallman lauds the detailed level of analysis. In particular, he suggests that such analysis is a royal replacement for the all too commonly used assumptions of fixed or simple trends in labor force participation rates.

POTENTIAL OUTPUT IN A RAPIDLY DEVELOPING ECONOMY: THE CASE OF CHINA AND A COMPARISON WITH THE UNITED STATES AND THE EUROPEAN UNION

China has experienced phenomenal economic growth during the past decade. In this study,

Jinghai Zheng, Angang Hu, and Arne Bigsten examine the sustainability of the growth process. Using the familiar growth accounting framework as their primary analytical tool, they address two major questions. First, to what extent has China's growth been extensive, that is, due primarily to increases in the quantities of inputs, versus intensive, that is, due to increases in total factor productivity? Second, if the growth is extensive, the authors conclude that growth cannot be sustained at its recent pace. The authors build a components-based estimate of potential output within the growth accounting framework and compare that measure with official statistics. They conclude that from 2002-08 the Chinese economy's output was above the measured level of potential—and hence not sustainable. In an innovative analysis, the authors measure the natural resources, or environmental, inputs to China's recent growth, concluding they have been important but that their contribution will diminish as China places new emphasis on the quality of its environment. In his commentary, Ziaodong Zhu notes that China's growth experience perhaps has been more subtle than the authors suggest. In aggregate data, he notes that labor force growth increased sharply during the first half of the reform period, 1978-90, but the investment rate increased little; as a result, the capital-to-output ratio fell. Labor force growth slowed sharply after 1990 and the investment rate increased, allowing steady growth in the capital-to-output ratio. In addition, the period since 1990 has been characterized by shrinkage of the state-owned sector and expansion of the private sector. At the start of the reform period in 1978, the capital-to-output ratio in the state-owned sector exceeded the ratio in the non-state-owned sector, while the marginal product of capital was similar in the two sectors. During the past decade, the state-owned sector experienced a nearly fourfold increase in its capital-to-labor ratio while its share of aggregate employment has increased and the capital-to-labor ratio in the non-state-owned sector approximately doubled. These differences in capital ratios and investment suggest that interpretations and projections of China's growth based on aggregate data may be hazardous.

ESTIMATING U.S. OUTPUT GROWTH WITH VINTAGE DATA IN A STATE-SPACE FRAMEWORK

In the conference's second paper based on vintage data, Richard Anderson and Charles Gascon explore estimating the "true" unobserved level of total output in the U.S. economy via a state-space model in which the true, or potential, output measure is unobserved. This analytical framework has been applied to U.K. data by staff of the Bank of England. Under certain assumptions, Monte Carlo simulations suggest that this framework can improve the accuracy of published estimates (relative to future revisions) by as much as 30 percent for as long as 11 periods. Real-time experiments show improvements closer to 10 percent, primarily during the first and second quarters of publication of revised data for a specific quarter. In his commentary, Dean Croushore explains that Anderson and Gascon's procedure may be defeated by the same aspect of vintage output data that has doomed earlier efforts: The data are subject to significant revisions—breaks—many years in the future, breaks that are difficult to predict in any filtering context. Using data drawn from the Federal Reserve Bank of St. Louis ALFRED database, Croushore explores both CBO projections of potential output and vintage data on output—the impacts of future unforecastable revisions (breaks) are difficult to overstate.

A PANEL DISCUSSION OF POLICYMAKERS

The conference concluded with three policymakers discussing their views and experience regarding the use of potential output in policymaking.

Carlos Araujo of the Central Bank of Brazil surveys the difficulty of using potential output for policy in a developing economy. Data revisions (including changes in methodology) are troublesome, and the length of the Brazilian time series is short. To maximize information extracted from the data, the Bank staff use a wide set of statistical filters and econometric modeling tools, despite

the efficacy of some filtering methods relying on statistical assumptions that are almost the opposite of others. In particular, various estimates are model dependent with disagreement regarding the appropriate model. Measures of potential output and its cousin, the output gap, are very important for policy but also must be recognized as imprecisely estimated.

Seppo Honkapohja of the Bank of Finland emphasizes that "there are two different concepts of the output gap and both are used in monetary policy analysis." The traditional concept, in his view, is the difference between a long-run trend and actual output; the more recent model-based concept is the difference between a model-based flexible price level of output and actual output. He notes that the two concepts are different and can behave in quite different ways through time—but both are used in policymaking. He cautions that model-based policy analysis, while attractive in its rigor and elegance, must be tempered by the shortcomings of the real world, including noisy data, broad issues of imperfect knowledge (including uncertainty regarding the correct model of the economy), and the effects of learning by economic agents.

James Bullard of the Federal Reserve Bank of St. Louis offered a challenge to modelers by arguing that any "trend" or potential output measure must be an integral part of the macro modeling itself. Bullard argues that such a viewpoint is essential throughout the "equilibrium" business cycle literature, including real business cycle, New Keynesian, and multisector growth models, because all equilibrium models are based on a long-run balanced growth path concept, which necessarily has implications for the allowable trends. Specifically, Bullard argues that the current practice in which data are "detrended" by atheoretic, univariate statistical methods is not defensible. Structural breaks are an important part of Bullard's analysis. He argues that empirical studies have shown convincingly that the economy's data-generating process experiences structural shifts. In such an economy, the macroeconomic concept of learnability is crucial—policymakers must recognize that economic agents *do* learn following a shift in trend and that optimal policy reactions are not invariant to the learning process.

What Do We Know (And Not Know) About Potential Output?

Susanto Basu and John G. Fernald

Potential output is an important concept in economics. Policymakers often use a one-sector neoclassical model to think about long-run growth, and they often assume that potential output is a smooth series in the short run—approximated by a medium- or long-run estimate. But in both the short and the long run, the one-sector model falls short empirically, reflecting the importance of rapid technological change in producing investment goods; and few, if any, modern macroeconomic models would imply that, at business cycle frequencies, potential output is a smooth series. Discussing these points allows the authors to discuss a range of other issues that are less well understood and where further research could be valuable. (JEL E32, O41, E60)

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 187-213.

The concept of potential output plays a central role in policy discussions. In the long run, faster growth in potential output leads to faster growth in actual output and, for given trends in population and the workforce, faster growth in income per capita. In the short run, policymakers need to assess the degree to which fluctuations in observed output reflect the economy's optimal response to shocks, as opposed to undesirable deviations from the time-varying optimal path of output.

To keep the discussion manageable, we confine our discussion of potential output to neoclassical growth models with exogenous technical progress in the short and the long run; we also focus exclusively on the United States. We make two main points. First, in both the short and the long run, rapid technological change in producing equipment investment goods is important. This rapid change in the production technology for investment goods implies that the two-sector neoclassical model—where one sector produces investment goods and the other produces consumption goods—provides a better benchmark for measuring potential output than the one-sector growth model. Second, in the short run, the measure of potential output that matters for policymakers is likely to fluctuate substantially over time. Neither macroeconomic theory nor existing empirical evidence suggests that potential output is a smooth series. Policymakers, however, often appear to assume that, even in the short run, potential output is well approximated by a smooth trend.[1] Our model and empirical work corroborate these two points and provide a framework to discuss other aspects of what we know, and do not know, about potential output.

As we begin, clear definitions are important to our discussion. "Potential output" is often used

[1] See, for example, Congressional Budget Office (CBO, 2001 and 2004) and Organisation for Economic Co-operation and Development (2008).

Susanto Basu is a professor in the department of economics at Boston College, a research associate of the National Bureau of Economic Research, and a visiting scholar at the Federal Reserve Bank of Boston. John G. Fernald is a vice president and economist at the Federal Reserve Bank of San Francisco. The authors thank Alessandro Barattieri and Kyle Matoba for outstanding research assistance and Jonas Fisher and Miles Kimball for helpful discussions and collaboration on related research. They also thank Bart Hobijn, Chad Jones, John Williams, Rody Manuelli, and conference participants for helpful discussions and comments.

© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

to describe related, but logically distinct, concepts. First, people often mean something akin to a "forecast" for output and its growth rate in the longer run (say, 10 years out). We will often refer to this first concept as a "steady-state measure," although a decade-long forecast can also incorporate transition dynamics toward the steady state.[2]

In the short run, however, a steady-state notion is less relevant for policymakers who wish to stabilize output or inflation at high frequencies. This leads to a second concept, explicit in New Keynesian dynamic stochastic general equilibrium (DSGE) models: Potential output is the rate of output the economy would have if there were no nominal rigidities but all other (real) frictions and shocks remained unchanged.[3] In a flexible price real business cycle model, where prices adjust instantaneously, potential output is equivalent to actual, equilibrium output. In contrast to the first definition of potential output as exclusively a long-term phenomenon, the second meaning defines it as relevant for the short run as well, when shocks push the economy temporarily away from steady state.

In New Keynesian models, where prices and/or wages might adjust slowly toward their long-run equilibrium values, actual output might well deviate from the short-term measure of potential output. In many of these models, the "output gap"—the difference between actual and potential output—is the key variable in determining the evolution of inflation. Thus, this second definition also corresponds to the older Keynesian notion that potential output is the "maximum production without inflationary pressure" (Okun, 1970, p. 133)—that is, the level of output at which there is no pressure for inflation to either increase or decrease. In most, if not all, macroeconomic models, the second (flexible price) definition converges in the long run to the first steady-state definition.

Yet a third definition considers potential output as the current *optimal* rate of output. With distortionary taxes and other market imperfections (such as monopolistic competition), neither steady-state output nor the flexible price equilibrium level of output needs to be optimal or efficient. Like the first two concepts, this third meaning is of interest to policymakers who might seek to improve the efficiency of the economy.[4] (However, decades of research on time inconsistency suggest that such policies should be implemented by fiscal or regulatory authorities, who can target the imperfections directly, but not by the central bank, which typically must take these imperfections as given. See, for example, the seminal paper by Kydland and Prescott, 1977.)

This article focuses on the first two definitions. The first part of our article focuses on long-term growth, which is clearly an issue of great importance for the economy, especially in discussions of fiscal policy. For example, whether promised entitlement spending is feasible depends almost entirely on long-run growth. We show that the predictions of two-sector models lead us to be more optimistic about the economy's long-run growth potential. This part of our article, which corresponds to the first definition of potential output, will thus be of interest to fiscal policymakers.

The second part of our article, of interest to monetary policymakers, focuses on a time-varying measure of potential output—the second usage above. Potential output plays a central, if often implicit, role in monetary policy decisions. The Federal Reserve has a dual mandate to pursue low and stable inflation and maximum sustainable employment. "Maximum sustainable employment" is usually interpreted to imply that the Federal Reserve should strive, subject to its other mandate, to stabilize the real economy around its flexible price equilibrium level—which itself is changing in response to real shocks—to avoid inefficient fluctuations in employment. In New Keynesian models, deviations of actual from potential output put pressure on inflation, so in

[2] In some models, transition dynamics can be very long-lived. For example, Jones (2002) interprets the past century as a time when growth in output per capita was relatively constant at a rate above steady state.

[3] See Woodford (2003) for the theory. Neiss and Nelson (2005) construct an output gap from a small, one-sector DSGE model.

[4] Justiniano and Primiceri (2008) define "potential output" as this third measure, with no market imperfections; they use the term "natural output" to mean our second, flexible-wage/price measure.

the simplest such models, output stabilization and inflation stabilization go hand in hand.

The first section of this article compares the steady-state implications of one- and two-sector neoclassical models with exogenous technological progress. That is, we focus on the long-run effects of given trends in technology, rather than trying to understand the sources of this technological progress.[5] Policymakers must understand the nature of technological progress to devise policies to promote long-run growth, but it is beyond the scope of our article. In the next section, we use the two-sector model to present a range of possible scenarios for long-term productivity growth and discuss some of the questions these different scenarios pose.

We then turn to short-term implications and ask whether it is plausible to think of potential output as a smooth process and compare the implications of a simple one-sector versus two-sector model. The subsequent section turns to the current situation (as of late 2008): How does short-run potential output growth compare with its steady-state level? This discussion suggests a number of additional issues that are unknown or difficult to quantify. The final section summarizes our findings and conclusions.

THE LONG RUN: WHAT SIMPLE MODEL MATCHES THE DATA?

A common, and fairly sensible, approach for estimating steady-state output growth is to estimate growth in full-employment labor productivity and then allow for demographics to determine the evolution of the labor force. This approach motivates this section's assessment of steady-state labor productivity growth.

We generally think that, in the long run, different forces explain labor productivity and total hours worked—technology along with induced capital deepening explains the former and demographics explains the latter. The assumption that labor productivity evolves separately from hours worked is motivated by the observation that labor productivity has risen dramatically over the past two centuries, whereas labor supply has changed by much less.[6] Even if productivity growth and labor supply are related in the long run, as suggested by Elsby and Shapiro (2008) and Jones (1995), the analysis that follows will capture the key properties of the endogenous response of capital deepening to technological change.

A reasonable way to estimate steady-state labor productivity growth is to estimate underlying technology growth and then use a model to calculate the implications for capital deepening. Let hats over a variable represent log changes. As a matter of identities, we can write output growth, $\hat{y}$, as labor-productivity growth plus growth in hours worked, $\hat{h}$:

$$\hat{y} = \left(\hat{y} - \hat{h}\right) + \hat{h}.$$

We focus here on full-employment labor productivity.

Suppose we define growth in total factor productivity (TFP), or the Solow residual, as

$$\widehat{tfp} = \hat{y} - \alpha\hat{k} - (1-\alpha)\hat{l},$$

where α is capital's share of income and $(1-\alpha)$ is labor's share of income. Defining

$$\hat{l} \equiv \hat{h} + \widehat{lq},$$

where $\widehat{lq}$ is labor "quality" (composition) growth,[7] we can rewrite output per hour growth as follows:

$$\left(\hat{y} - \hat{h}\right) = \widehat{tfp} + \alpha\left(\hat{k} - \hat{l}\right) + \widehat{lq}. \tag{1}$$

As an identity, growth in output per hour worked reflects TFP growth; the contribution of

[5] Of course, total factor productivity (TFP) can change for reasons broader than technological change alone; improved institutions, deregulation, and less distortionary taxes are only some of the reasons. We believe, however, that long-run trends in TFP in developed countries like the United States are driven primarily by technological change. For evidence supporting this view, see Basu and Fernald (2002).

[6] King, Plosser, and Rebelo (1988) suggest a first approximation should model hours per capita as independent of the level of technology and provide necessary and sufficient conditions on the utility function for this result to hold. Basu and Kimball (2002) show that the particular non-separability between consumption and hours worked that is generally implied by the King-Plosser-Rebelo utility function helps explain the evolution of consumption in postwar U.S. data and resolves several consumption puzzles.

[7] See footnote 7 on p. 190.

capital deepening, defined as $\alpha(\hat{k} - \hat{l})$; and increases in labor quality. Economic models suggest mappings between fundamentals and the terms in this identity that are sometimes trivial and sometimes not.

The One-Sector Model

Perhaps the simplest model that could reasonably be applied to the long-run data is the one-sector neoclassical growth model. Technological progress and labor force growth are exogenous and capital deepening is endogenous.

We can derive the key implications from the textbook Solow version of the model. Consider an aggregate production function $Y = K^{\alpha}(AL)^{1-\alpha}$, where technology A grows at rate g and labor input L (which captures both raw hours, H, and labor quality, LQ—henceforth, we do not generally differentiate between the two) grows at rate n. Expressing all variables in terms of "effective labor," AL, yields

$$y = k^{\alpha}, \tag{2}$$

where $y = Y/AL$ and $k = K/AL$.

Capital accumulation takes place according to the perpetual-inventory formula. If s is the saving rate, so that sy is investment per effective worker, then in steady state

$$sy = (n + \delta + g)k. \tag{3}$$

Because of diminishing returns to capital, the economy converges to a steady state where y and k are constant. At that point, investment per effective worker is just enough to offset the effects of depreciation, population growth, and technological change on capital per effective worker. In steady state, the unscaled levels of Y and K grow at the rate $g + n$; capital deepening, K/L, grows at rate g. Labor productivity Y/L (i.e., output per unit of labor input) also grows at rate g.

From the production function, measured TFP growth is related to labor-augmenting technology growth by

$$\widehat{tfp} = \hat{Y} - \alpha\hat{K} - (1-\alpha)\hat{L} = (1-\alpha)g.$$

The model maps directly to equation (1). In particular, the endogenous contribution of capital deepening to labor-productivity growth is

$$\alpha g = \alpha \cdot \widehat{tfp} / (1-\alpha).$$

Output per unit of labor input grows at rate g, which equals the sum of standard TFP growth, $(1 - \alpha)g$, and induced capital deepening, αg.

Table 1 shows how this model performs relative to the data. It uses the multifactor productivity release from the Bureau of Labor Statistics (BLS), which provides data for TFP growth as well as capital deepening for the U.S. business economy. These data are shown in the first two columns. Note that in the model above, standard TFP growth reflects technology alone. In practice, a large segment of the literature suggests reasons why nontechnological factors might affect measured TFP growth. For example, there are hard-to-measure short-run movements in labor effort and capital's workweek, which cause measured (although not actual) TFP to fluctuate in the short run. Nonconstant returns to scale and markups also interfere with the mapping from technological change to measured aggregate TFP. But the deviations between technology and measured TFP are likely to be more important in the short run than in the long run, consistent with the findings of Basu, Fernald, and Kimball (2006) and Basu et al. (2008). Hence, for these longer-term comparisons, we assume average TFP growth reflects average technology growth. Column 3 shows the predictions of the one-sector neoclassical model for $\alpha = 0.32$ (the average value in the BLS multifactor dataset).

---

[7] Labor quality/composition reflects the mix of hours across workers with different levels of education, experience, and so forth. For the purposes of this discussion, which so far has focused on definitions, suppose there were J types of workers with factor shares of income β_j, where

$$\sum_j \beta_j = (1-\alpha).$$

Then a reasonable definition of TFP would be

$$\widehat{tfp} = \hat{y} - \alpha\hat{k} - \sum_j \beta_j \hat{h}_j.$$

Growth accounting as done by the Bureau of Labor Statistics or by Dale Jorgenson and his collaborators (see, for example, Jorgenson, Gollop, and Fraumeni, 1987) defines

$$\hat{l} = \sum_j \beta_j \hat{h}_j \big/ (1-\alpha), \quad \hat{h} = d\ln\sum_j H_j, \quad \text{and} \quad \hat{q} = \hat{l} - \hat{h}.$$

Table 1

One-Sector Growth Model Predictions for the U.S. Business Sector

Period	Total TFP	Actual capital deepening contribution	Predicted capital deepening contribution in one-sector model
1948-2007	1.39	0.76	0.65
1948-1973	2.17	0.85	1.02
1973-1995	0.52	0.62	0.25
1995-2007	1.34	0.84	0.63
1995-2000	1.29	1.01	0.61
2000-2007	1.37	0.72	0.65

NOTE: Data for columns 1 and 2 are business sector estimates from the BLS multifactor productivity database (downloaded via Haver on August 19, 2008). Capital and labor are adjusted for changes in composition. Actual capital deepening is $\alpha(\hat{k} - \hat{l})$, and predicted capital deepening is $\alpha \cdot \widehat{tfp} / (1-\alpha)$.

A comparison of columns 2 and 3 shows the model does not perform particularly well. It slightly underestimates the contribution of capital deepening over the entire 1948-2007 period, but it does a particularly poor job of matching the low-frequency variation in that contribution. In particular, it somewhat overpredicts capital deepening for the pre-1973 period but substantially underpredicts for the 1973-95 period. That is, given the slowdown in TFP growth, the model predicts a much larger slowdown in the contribution of capital deepening.[8]

One way to visualize the problem with the one-sector model is to observe that the model predicts a constant capital-to-output ratio in steady state—in contrast to the data. Figure 1 shows the sharp rise in the business sector capital-to-output ratio since the mid-1960s.

The Two-Sector Model: A Better Match

A growing literature on investment-specific technical change suggests an easy fix for this failure: Capital deepening does not depend on *overall* TFP but on TFP in the investment sector. A key motivation for this body of literature is the price of business investment goods, especially equipment and software, relative to the price of other goods (such as consumption). The relative price of investment and its main components are shown in Figure 2.

Why do we see this steady relative price decline? The most natural interpretation is that there is a more rapid pace of technological change in producing investment goods (especially high-tech equipment).[9]

To realize the implications of a two-sector model, consider a simple two-sector Solow-type model, where s is the share of nominal output that is invested each period.[10] One sector produces investment goods, I, that are used to create capital; the other sector produces consumption goods, C. The two sectors use the same Cobb-Douglas production function but with potentially different technology levels:

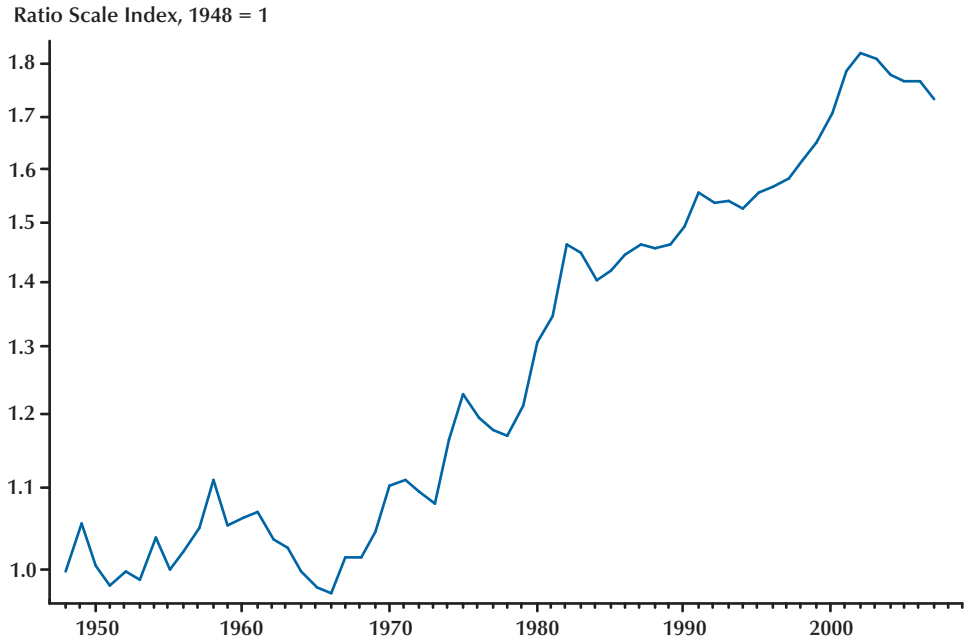
[8] Note that output per unit of quality-adjusted labor is the sum of TFP plus the capital deepening contribution, which in the business sector averaged 1.39 + 0.76 = 2.15 percent per year over the full sample. More commonly, labor productivity is reported as output per hour worked. Over the sample, labor quality in the BLS multifactor productivity dataset rose 0.36 percent per year, so output per *hour* rose 2.51 percent per year.

[9] On the growth accounting side, see, for example, Jorgenson (2001) or Oliner and Sichel (2000); see also Greenwood, Hercowitz, and Krusell (1997).

[10] This model is a fixed–saving rate version of the two-sector neoclassical growth model in Whelan (2003) and is isomorphic to the one in Greenwood, Hercowitz, and Krusell (1997), who choose a different normalization of the two technology shocks in their model.

Figure 1

Capital-to-Output Ratio in the United States (equipment and structures)



SOURCE: BLS multisector productivity database. Equipment and structures (i.e., fixed reproducible tangible capital) is calculated as a Tornquist index of the two categories. Standard Industrial Classification data (from www.bls.gov/mfp/historicalsic.htm) are spliced to North American Industry Classification System data (from www.bls.gov/mfp/mprdload.htm) starting at 1988 (data downloaded October 13, 2008).

$$I = K_I^{\alpha}\left(A_I L_I\right)^{1-\alpha}$$
$$C = QK_C^{\alpha}\left(A_I L_C\right)^{1-\alpha}.$$

In the consumption equation, we have implicitly defined labor-augmenting technological change as $A_C = Q^{1/(1-\alpha)}A_I$ to decompose consumption technology into the product of investment technology, A_I, and a "consumption-specific" piece, $Q^{1/(1-\alpha)}$. Let investment technology, A_I, grow at rate g_I and the consumption-specific piece, Q, grow at rate q. Perfect competition and cost minimization imply that price equals marginal cost. If the sectors face the same factor prices (and the same rate of indirect business taxes), then

$$\frac{P_I}{P_C} = \frac{MC^C}{MC^I} = Q.$$

The sectors also choose to produce with the same capital-to-labor ratios, implying that $K_I/A_I L_I = K_C/A_I L_C = K/A_I L$. We can then write the production functions as

$$I = A_I L_I \left(K/A_I L\right)^{\alpha}$$
$$C = QA_I L_C \left(K/A_I L\right)^{\alpha}.$$

We can now write the economy's budget constraint in a simple manner:

$$Y^{\text{Inv. Units}} \equiv \left[I + C/Q\right] = A_I\left(L_I + L_C\right)\left(K/A_I L\right)^{\alpha},$$

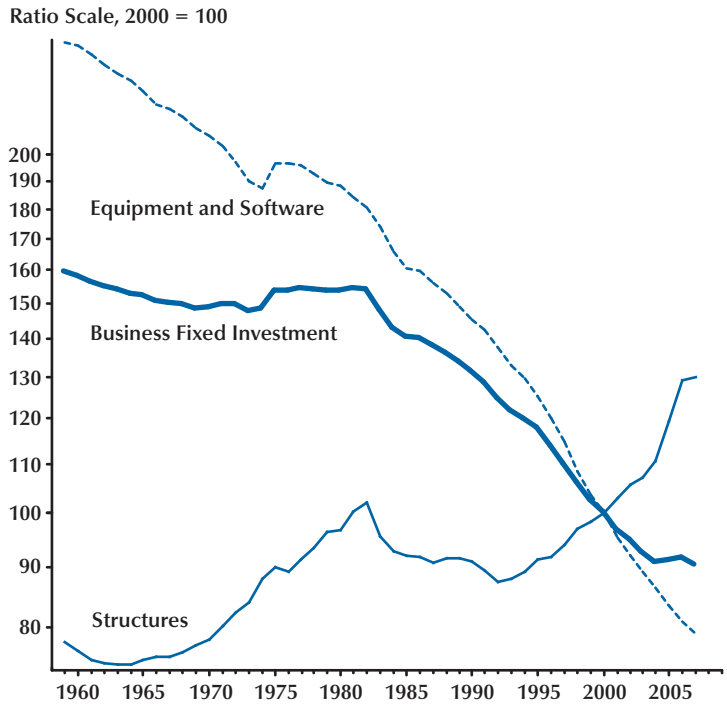
(4) or $y^{\text{Inv. Units}} = k^{\alpha}$, where

$$y^{\text{Inv. Units}} = Y^{\text{Inv. Units}}/A_I L \text{ and } k = K/A_I L.$$

"Output" here is expressed in investment units, and "effective labor" is in terms of technology in the *investment* sector. The economy mechanically invests a share s of nominal invest-

Figure 2

Price of Business Fixed Investment Relative to Other Goods and Services



NOTE: "Other goods and services" constitutes business GDP less business fixed investment.
SOURCE: Bureau of Economic Analysis and authors' calculations.

ment, which implies that investment per effective unit of labor is $i = s \cdot y^{\text{Inv. Units}}$.[11] Capital accumulation then takes the same form as in the one-sector model, except that it is only growth in investment technology, g_I, that matters. In particular, in steady state,

$$sy^{\text{Inv. Units}} = (n + \delta + g_I)k. \tag{5}$$

The production function (4) and capital-accumulation equation (5) correspond exactly to their one-sector counterparts. Hence, the dynamics of capital in this model reflect technology in the investment sector alone. In steady state, capital per unit of labor, K/L, grows at rate g_I, so the contribution of capital deepening to labor-productivity growth from equation (1) is

$$\alpha g_I = \alpha \cdot \widehat{tfp_I} / (1 - \alpha).$$

Consumption technology in this model is "neutral" in that it does not affect investment or capital accumulation; the same result carries over to the Ramsey version of this model, with or without variable labor supply. (Basu et al., 2008, discuss the idea of consumption-technology neutrality in greater detail.[12])

To apply this model to the data, we need to decompose aggregate TFP growth (calculated from

[11] $s \cdot y^{\text{Inv. Units}} = \left[P_I I / (P_I I + P_C C)\right]\left[(I + P_C C / P_I) / A_I L\right] = I / A_I L.$

[12] Note also that output in investment units is not equal to chain output in the national accounts. Chain gross domestic product (GDP) is

$$\hat{Y} = s\hat{I} + (1 - s)\hat{C}.$$

In contrast, in this model $\widehat{Y^{\text{Inv. Units}}} = s\hat{I} + (1 - s)\hat{C} - (1 - s)q.$

Hence, $\hat{Y} = \widehat{Y^{\text{Inv. Units}}} + (1 - s)q.$

chained output) into its consumption and investment components. Given the conditions so far, the following two equations hold:

$$\widehat{tfp} = s \cdot \widehat{tfp}_I + (1-s)\widehat{tfp}_C,$$
$$\widehat{P_C} - \widehat{P_I} = \widehat{tfp}_C - \widehat{tfp}_I.$$

These are two equations in two unknowns—

$$\widehat{tfp}_I \text{ and } \widehat{tfp}_C.$$

Hence, they allow us to decompose aggregate TFP growth into investment and consumption TFP growth.[13]

Table 2 shows that the two-sector growth model does, in fact, fit the data better. All derivations are done assuming an investment share of 0.15, about equal to the nominal value of business fixed investment relative to the value of business output.

For the 1948-73 and 1973-95 periods, a comparison of columns 5 and 6 indicates that the model fits quite well—and much better than the one-sector model. The improved fit reflects that although overall TFP growth slowed very sharply, investment TFP growth (column 3) slowed much less. Hence, the slowdown in capital deepening was much smaller.

The steady-state predictions work less well for the periods after 1995, when actual capital deepening fell short of the steady-state prediction for capital deepening. During these periods, not only did overall TFP accelerate, but the relative price decline in column 2 also accelerated. Hence, implied investment TFP accelerated markedly (as did other TFP). Of course, the transition dynamics imply that capital deepening converges only slowly to the new steady state, and a decade is a relatively short time. (In addition, the pace of investment-sector TFP was particularly rapid in the late 1990s and has slowed somewhat in the 2000s.) So the more important point is that, qualitatively, the model works in the right direction even over this relatively short period.

Despite these uncertainties, a bottom-line comparison of the one- and two-sector models is of interest. Suppose that the 1995-2007 rates of TFP growth continue to hold in both sectors (a big "if" discussed in the next section). Suppose also that the two-sector model fits well going forward, as it did in the 1948-95 period. Then we would project that future output per hour (like output per quality-adjusted unit of labor, shown in Tables 1 and 2) will grow on average about 0.75 percentage points per year faster than the one-sector model would predict (1.38 versus 0.63), as a result of greater capital deepening. The difference is clearly substantial: It is a significant fraction of the average 2.15 percent growth rate in output per unit of labor (and 2.5 percent growth rate of output per hour) over the 1948-2007 period.

PROJECTING THE FUTURE

Forecasters, policymakers, and a number of academics regularly make "structured guesses" about the likely path of future growth.[14] Not surprisingly, the usual approach is to assume that the future will look something like the past—but the challenge is to decide which parts of the past to include and which to downplay.

In making such predictions, economists often project average TFP growth for the economy as a whole. However, viewed through the lens of the two-sector model, one needs to make separate projections for TFP growth in both the investment and non-investment sectors. We consider three growth scenarios: low, medium, and high (Table 3).

Consider the medium scenario, which has output per hour growing at 2.3 (last column). Investment TFP is a bit slower than its average in the post-2000 period, reflecting that investment TFP has generally slowed since the burst of the late 1990s. Other TFP slows to its rate in

[13] The calculations below use the official price deflators from the national accounts. Gordon (1990) argues that many equipment deflators are not sufficiently adjusted for quality improvements over time. Much of the macroeconomic literature since then has used the Gordon deflators (possibly extrapolated, as in Cummins and Violante, 2002). Of course, as Whelan (2003) points out, much of the discussion of biases in the consumer price index involves service prices, which also miss many quality improvements.

[14] Oliner and Sichel (2002) use the phrase "structured guesses." In addition to Oliner and Sichel, recent high-profile examples of projections have come from Jorgenson, Ho, and Stiroh (2008) and Gordon (2006). The CBO and the Council of Economic Advisers regularly include longer-run projections of potential output.

Table 2

Two-Sector Growth Model Predictions for the U.S. Business Sector

Period	Total TFP	Relative price of business fixed investment to other goods and services	Investment TFP	Other TFP	Actual capital deepening contribution	Predicted capital deepening contribution in two-sector model
1948-2007	1.39	–0.61	1.91	1.29	0.76	0.90
1948-1973	2.17	0.33	1.89	2.22	0.85	0.89
1973-1995	0.52	–1.02	1.39	0.37	0.62	0.66
1995-2007	1.34	–1.90	2.94	1.04	0.84	1.38
1995-2000	1.29	–2.93	3.78	0.85	1.01	1.78
2000-2007	1.37	–1.17	2.36	1.20	0.72	1.11
2004:Q4–2006:Q4	0.21	0.29	–0.04	0.25	—	—
2006:Q4–2008:Q3	0.98	–1.12	1.94	0.82	—	—

NOTE: "Other goods and services" constitutes business GDP less business fixed investment. Capital and labor are adjusted for changes in composition. Actual capital deepening is $\alpha(\hat{k} - \hat{l})$, and predicted capital deepening is $\alpha \cdot \widehat{tfp} / (1 - \alpha)$.

SOURCE: BLS multifactor productivity dataset, Bureau of Economic Analysis relative-price data, and authors' calculations. The final two rows reflect quarterly estimates from Fernald (2008); because of the very short sample periods, we do not show steady-state predictions.

Table 3

A Range of Estimates for Steady-State Labor Productivity Growth

Growth scenario	Investment TFP	Other TFP	Overall TFP	Capital deepening contribution	Labor productivity	Output per hour worked
Low	1.00	0.70	0.7	0.5	1.2	1.5
Medium	2.00	0.82	1.0	0.9	2.0	2.3
High	2.50	1.10	1.3	1.2	2.5	2.8

NOTE: Calculations assume an investment share of output of 0.15 and a capital share in production, α, of 0.32. Column 3 (Overall TFP) is an output-share-weighted average of columns 1 and 2. Column 4 is column 1 multiplied by $\alpha/(1-\alpha)$. Column 5 is output per unit of composition-adjusted labor input and is the sum of columns 3 and 4. Column 6 adds an assumed growth rate of labor quality/composition of 0.3 percent per year, and therefore equals column 5 plus 0.3 percent.

the second half of the 1990s, reflecting an assumption that the experience of the early 2000s is unlikely to persist.

Productivity growth averaging about 2.25 percent is close to a consensus forecast. For example, in the first quarter of 2008, the median estimate in the Survey of Professional Forecasters (SPF, 2008) was for 2 percent labor-productivity growth over the next 10 years (and 2.75 percent gross domestic product [GDP] growth). In September 2008, the Congressional Budget Office estimated that labor productivity (in the nonfarm business sector) would grow at an average rate of about 2.2 percent between 2008 and 2018.[15]

As Table 3 clearly shows, however, small and plausible changes in assumptions—well within the range of recent experience—can make a large difference for steady-state growth projections. As a result, a wide range of plausible outcomes exists. In the SPF, the standard deviation across the 39 respondents for productivity growth over the next 10 years was about 0.4 percent—with a range of 0.9 to 3.0 percent. Indeed, the current median estimate of 2.0 percent is down from an estimate of 2.5 percent in 2005, but remains much higher than the one-year estimate of only 1.3 percent in 1997.[16]

[15] Calculated from data in CBO (2008).

[16] The SPF has been asking about long-run projections in the first quarter of each year since 1992. The data are available at www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/data-files/PROD10/.

The two-sector model suggests several key questions in making long-run projections. First, what will be the pace of technical progress in producing information technology (IT) and, more broadly, equipment goods? For example, for hardware, Moore's law—that semiconductor capacity doubles approximately every two years—provides plausible bounds. For software, however, we really have very little firm ground for speculation.

Second, how elastic is the demand for IT? The previous discussion of the two-sector model assumed that the investment share was constant at 0.15. But an important part of the price decline reflected that IT, for which prices have been falling rapidly, is becoming an increasing share of total business fixed investment. At some point, a constant share is a reasonable assumption and consistent with a balanced growth path. Yet over the next few decades, very different paths are possible. Technology optimists (such as DeLong, 2002) think that the elasticity of demand for IT exceeds unity, so that demand will rise even faster than prices fall. They think that firms and individuals will find many new uses for computers, semiconductors, and, indeed, information, as these commodities get cheaper and cheaper. By contrast, technology pessimists (such as Gordon, 2000) think that the greatest contribution of the IT revolution is in the past rather than the future. For example, firms may decide they will not need much more computing power in the future, so that as prices continue to fall, the nominal share

of expenditure on IT will also fall. For example, new and faster computers might offer few advantages for word processing relative to existing computers, so the replacement cycle might become longer.

Third, what will happen to TFP in the non-IT-producing sectors? The range of uncertainty here is very large—larger, arguably, than for the first two questions. The general-purpose-technology nature of computing suggests that faster computers and better ability to manage and manipulate information might well lead to TFP improvements in computer-using sectors.[17] For example, many important management innovations, such as the Wal-Mart business model or the widespread diffusion of warehouse automation, are made possible by cheap computing power. Productivity in research and development may also rise more directly; auto parts manufacturers, for example, can design new products on a computer rather than building physical prototype models. That is, computers may lower the cost and raise the returns to research and development

In addition, are these sorts of TFP spillovers from IT to non-IT sectors best considered as growth effects or level effects? For example, the "Wal-Martization" of retailing raises productivity levels (as more-efficient producers expand and less-efficient producers contract) but it does not necessarily boost long-run growth.

Fourth, the effects noted previously might well depend on labor market skills. Many endogenous growth models incorporate a key role for human capital, which is surely a key input into the innovation process—whether reflected in formal research and development or in management reorganizations. Beaudry, Doms, and Lewis (2006) find evidence that the intensity of personal computers use across U.S. cities is closely related to education levels in those cities.

We hope we have convinced readers that it is important to take a two-sector approach to estimating the time path of long-run output. But as this (non-exhaustive) discussion demonstrates, knowing the correct framework for analysis is only one of many inputs to projecting potential output correctly. Much still remains unknown about potential output, even along a steady-state growth path. The biggest problem is the lack of knowledge about the deep sources of TFP growth.

[17] See, for example, Basu et al. (2003) for an interpretation of the broad-based TFP acceleration in terms of intangible organizational capital associated with using computers. Of course, an intangible-capital story suggests that the measured share of capital is too low, and that measured capital is only a subset of all capital—so the model and calibration in the earlier section are incomplete.

SHORT-RUN CONSIDERATIONS

General Issues in Defining and Estimating Short-Run Potential Output

Traditionally, macroeconomists have taken the view expressed in Solow (1997) that, in the long-run, a growth model such as the ones described previously explains the economy's long-run behavior. Factor supplies and technology determine output, with little role for "demand" shocks. However, the short run was viewed very differently, when as Solow (1997) put it, "…fluctuations are predominantly driven by aggregate demand impulses" (p. 230).

Solow (1997) recognizes that real business cycle theories take a different view, providing a more unified vision of long-run growth and short-run fluctuations than traditional Keynesian views did. Early real business cycle models, in particular, emphasized the role of high-frequency technology shocks. These models are also capable of generating fluctuations in response to nontechnological "demand" shocks, such as government spending. Since early real business cycle models typically do not incorporate distortions, they provide examples in which fluctuations driven by government spending or other impulses could well be optimal (taking the shocks themselves as given). Nevertheless, traditional Keynesian analyses often presumed that potential output was a smooth trend, so that any fluctuations were necessarily suboptimal (regardless of whether policy could do anything about them).

Fully specified New Keynesian models provide a way to think formally about the sources of business cycle fluctuations. These models are generally founded on a real business cycle model, albeit one with real distortions, such as firms

having monopoly power. Because of sticky wages and/or prices, purely nominal shocks, such as monetary policy shocks, can affect real outcomes. The nominal rigidities also affect how the economy responds to real shocks, whether to technology, preferences, or government spending. Short-run potential output is naturally defined as the rate of output the economy would have if there were no nominal rigidities, that is, by the responses in the real business cycle model underlying the sticky price model.[18] This is our approach to producing a time series of potential output fluctuations in the short run.

In New Keynesian models, where prices and/or wages might adjust slowly toward their long-run equilibrium values, actual output might well deviate from this short-term measure of potential output. In many of these models, the "output gap"—the difference between actual and potential output—is the key variable in determining the evolution of inflation. Kuttner (1994) and Laubach and Williams (2003) use this intuition to estimate the output gap as an unobserved component in a Phillips curve relationship. They find fairly substantial time variation in potential output.

In the context of New Keynesian DSGE models, is there any reason to think that potential output is a smooth series? At a minimum, a low variance of aggregate technology shocks as well as inelastic labor supply is needed. Rotemberg (2002), for example, suggests that because of slow diffusion of technology across producers, stochastic technological improvements might drive long-run growth without being an important factor at business cycle frequencies.[19]

Nevertheless, although a priori one might believe that technology changes only smoothly over time, there is scant evidence to support this position. Basu, Fernald, and Kimball (2006) control econometrically for nontechnological factors affecting the Solow residual—nonconstant returns to scale, variations in labor effort and capital's workweek, and various reallocation effects—and still find a "purified technology" residual that is highly variable. Alexopoulos (2006) uses publications of technical books as a proxy for unobserved technical change and finds that this series is not only highly volatile, but explains a substantial fraction of GDP and TFP. Finally, variance decompositions often suggest that innovations to technology explain a substantial share of the variance of output and inputs at business cycle frequencies; see Basu, Fernald, and Kimball (2006) and Fisher (2006).

When producing a time series of short-run potential output, it is necessary not only to know "the" correct model of the economy, but also the series of historical shocks that have affected the economy. One approach is to specify a model, which is often complex, and then use Bayesian methods to estimate the model parameters on the data. As a by-product, the model estimates the time series of all the shocks that the model allows.[20] Because DSGE models are "structural" in the sense of Lucas's (1976) critique, one can perform counterfactual simulations—for example, by turning off nominal rigidities and using the estimated model and shocks to create a time series of flexible price potential output.

We do not use this approach because we are not sure that Bayesian estimation of DSGE models always uses reliable schemes to identify the relevant shocks. The full-information approach of these models is, of course, preferable in an efficiency sense—if one is sure that one has specified

[18] See Woodford (2003). There is a subtle issue in defining flexible price potential output when the time path of actual output may be influenced by nominal rigidities. In theory, the flexible price output series should be a purely forward-looking construct, which is generated by "turning off" all nominal rigidities in the model, but starting from current values of all state variables, including the capital stock. Of course, the current value of the capital stock might be different from what it would have been in a flexible price model with the same history of shocks because nominal rigidities operated in the past. Thus, in principle, the potential-output series should be generated by initializing a flexible price model every period, rather than taking an alternative time-series history from the flexible price model hit by the same sequence of real shocks. We do the latter rather than the former because we believe that nominal rigidities cause only small deviations in the capital stock, but it is possible that the resulting error in our potential-output series might actually be important.

[19] A recent paper by Justiniano and Primiceri (2008) estimates both simple and complex New Keynesian models and finds that most of the volatility in the flexible-wage/price economy reflects extreme volatility in markup shocks. They still estimate that there is considerable quarter-to-quarter volatility in technology, so that even if the only shocks were technology shocks, their flexible price measure of output would also have considerable volatility from one quarter to the next.

[20] See Smets and Wouters (2007).

the correct structural model of the economy with all its frictions. We prefer to use limited-information methods to estimate the key shocks—technology shocks, in our case—and then feed them into small, plausibly calibrated models of fluctuations. At worst, our method should provide a robust, albeit inefficient, method of assessing some of the key findings of DSGE models estimated using Bayesian methods.

We believe that our method of estimating the key shocks is both more transparent in its identification and robust in its method because it does not rely on specifying correctly the full model of the economy, but only small pieces of such a model. As in the case of the Basu, Fernald, and Kimball (2006) procedure underlying our shock series, we specify only production functions and costs of varying factor utilization and assume that firms minimize costs—all standard elements of current "medium-scale" DSGE models. Furthermore, we assume that true technology shocks are orthogonal to other structural shocks, such as monetary policy shocks, which can therefore be used as instruments for estimation. Finally, because we do not have the overhead of specifying and estimating a complete structural general equilibrium model, we are able to model the production side of the economy in greater detail. Rather than assuming that an aggregate production function exists, we estimate industry-level production functions and aggregate technology shocks from these more disaggregated estimates. Basu and Fernald (1997) argue that this approach is preferable in principle and solves a number of puzzles in recent production-function estimation in practice.

We use time series of "purified" technology shocks, similar to those presented in Basu, Fernald, and Kimball (2006) and Basu et al. (2008). However, these series are at an annual frequency. Fernald (2008) applies the methods in these articles to quarterly data and produces higher-frequency estimates of technology shocks. Fernald estimates utilization-adjusted measures of TFP for the aggregate economy, as well as for the investment and consumption sector. In brief, aggregate TFP is measured using data from the BLS quarterly labor productivity data, combined with capital-service data estimated from detailed quarterly investment data. Labor quality and factor shares are interpolated from the BLS multifactor-productivity dataset. The relative price of investment goods is used to decompose aggregate TFP into investment and consumption components, using the (often-used) assumption that relative prices reflect relative TFPs. The utilization adjustment follows Basu, Fernald, and Kimball (2006), who use hours per worker as a proxy for utilization change (with an econometrically estimated coefficient) at an industry level. The input-output matrix was used to aggregate industry utilization change into investment and consumption utilization change, following Basu et al. (2008).[21]

To produce our estimated potential output series, we feed the technology shocks estimated by Fernald (2008) into simple one- and two-sector models of fluctuations (see the appendix). Technology shocks shift the production function directly, even if they are not amplified by changes in labor supply in response to variations in wages and interest rates. If labor supply is elastic, then a fortiori the changes in potential output will be more variable for any given series of technology shocks.

Elastic labor supply also allows nontechnology shocks to move short-run, flexible price output discontinuously. Shocks to government spending, even if financed by lump-sum taxes, cause changes in labor supply via a wealth effect. Shocks to distortionary tax rates on labor income shift labor demand and generally cause labor input, and hence output, to change. Shocks to the preference for consumption relative to leisure can also cause changes in output and its components.

The importance of all of these shocks for movements in flexible price potential output depends crucially on the size of the Frisch (wealth-constant) elasticity of labor supply. Unfortunately, this is one of the parameters in economics whose value is most controversial, at least at an aggregate level. Most macroeconomists assume values between 1 and 4 for this crucial

[21] Because of a lack of data at a quarterly frequency, Fernald (2008) does not correct for deviations from constant returns or for heterogeneity across industries in returns to scale—issues that Basu, Fernald, and Kimball (2006) argue are important.

parameter, but not for particularly strong reasons.[22] On the other hand, Card (1994) reviews both microeconomic and aggregative evidence and concludes there is little evidence in favor of a nonzero Frisch elasticity of labor supply. The canonical models of Hansen (1985) and Rogerson (1988) attempt to bridge the macro-micro divide. However, Mulligan (2001) argues that the strong implication of these models, an infinite aggregate labor supply elasticity, depends crucially on the assumption that workers are homogeneous and can easily disappear when one allows for heterogeneity in worker preferences.

We do not model real, nontechnological shocks to the economy in creating our series on potential output. Our decision is partly due to uncertainty over the correct value of the aggregate Frisch labor supply elasticity, which as discussed previously is crucial for calibrating the importance of such shocks. We also make this decision because in our judgment there is even less consensus in the literature over identifying true innovations to fiscal policy or to preferences than there is on identifying technology shocks. Our decision to ignore nontechnological real shocks clearly has the potential to bias our series on potential output, and depending on the values of key parameters, this bias could be significant.

One-Sector versus Two-Sector Models

In the canonical New Keynesian Phillips curve, derived with Calvo price setting and flexible wages, inflation today depends on expected inflation tomorrow, as well as on the gap between actual output and the level of output that would occur with flexible prices.

To assess how potential and actual output respond in the short run in a one- versus two-sector model, we used a very simple two-sector New Keynesian model (see the appendix). As in the long-run model, we assume that investment and consumption production uses a Cobb-Douglas technology with the same factor shares but with a (potentially) different multiplicative technology parameter. To keep things simple, factors are completely mobile, so that a one-sector model is the special case when the same technology shock hits both sectors.

We simulated the one- and two-sector models using the utilization-adjusted technology shocks estimated in Fernald (2008). Table 4 shows standard deviations of selected variables in flexible and sticky price versions of the one- and two-sector models, along with actual data for the U.S. economy.

The model does a reasonable job of approximating the variation in actual data, considering how simple it is and that only technology shocks are included. Investment in the data is slightly less volatile than either in the sticky price model or the two-sector flexible price model. This is not surprising, given that the model does not have any adjustment costs or other mechanisms to smooth out investment. Consumption, labor, and output in the data are more volatile than in the models.[23] Additional shocks (e.g., to government spending, monetary policy, or preferences) would presumably add volatility to model simulations.

An important observation from Table 4 is that potential output—the flexible price simulations, in either the one- or two-sector variants—is highly variable, roughly as variable as sticky price output. The short-run variability of potential output in New Keynesian models has been emphasized by Neiss and Nelson (2005) and Edge, Kiley, and Laforte (2007).

These models, with the shocks we have added, show a very high correlation of flexible and sticky price output. In the two-sector case, the correlation is 0.91. Nevertheless, the implied output gap (shown in the penultimate line of Table 4 as the difference between output in the flexible and sticky price cases) is more volatile than would be implied if potential output were estimated with the one-sector model (the final line).

[22] In many cases, it is simply because macro models do not "work"—that is, display sufficient amplification of shocks—for smaller values of the Frisch labor supply elasticity. In other cases, values like 4 are rationalized by assuming, without independent evidence, that the representative consumer's utility from leisure takes the logarithmic form. However, this restriction is not imposed by the King-Plosser-Rebelo (1988) utility function, which guarantees balanced growth for any value of the Frisch elasticity.

[23] The relative volatility of consumption is not that surprising, because the models do not have consumer durables and we have not yet analyzed consumption of nondurables and services in the actual data.

Table 4

Standard Deviations, Model Simulations, and Data

Variable	Investment	Consumption	Labor	Output
One-sector, flexible price	4.40	0.81	0.47	1.52
Two-sector, flexible price	6.28	0.89	0.73	1.66
One-sector, sticky price	4.82	0.84	0.64	1.60
Two-sector, sticky price	5.52	0.87	0.85	1.68
Data	4.54	1.12	1.14	1.95
Output gap (two-sector sticky price less two-sector flexible price)	5.78	0.59	0.96	0.72
"One-sector" estimated gap (two-sector sticky price less one-sector flexible price)	2.55	0.18	0.59	0.41

NOTE: Model simulations use utilization-adjusted TFP shocks from Fernald (2008). Two-sector simulations use estimated quarterly consumption and investment technology; one-sector simulations use the same aggregate shock (a share-weighted average of the two sectoral shocks) in both sectors. All variables are filtered with the Christiano-Fitzgerald bandpass filter to extract variation between 6 and 32 quarters.

Figure 3 shows that the assumption that potential output has no business cycle variation—which is tantamount to using (Hodrick-Prescott–filtered) sticky price output itself as a proxy for the output gap—would overestimate the variation in the output gap. This would not matter too much if the output gap were perfectly correlated with sticky price output itself—then, at least, the sign, if not the magnitude, would be correct. However, as the figure shows, the "true" two-sector output gap in the model (two-sector sticky price output less two-sector flexible price output) is imperfectly correlated with sticky price output—indeed, the correlation is only 0.25. So in this model, policymakers could easily be misled by focusing solely on output fluctuations rather than the output gap.

Implications for Stabilization Policy

If potential output fluctuates substantially over time, then this has potential implications for the desirability of stabilization policy. In particular, policymakers should be focused only on stabilizing *undesirable* fluctuations.

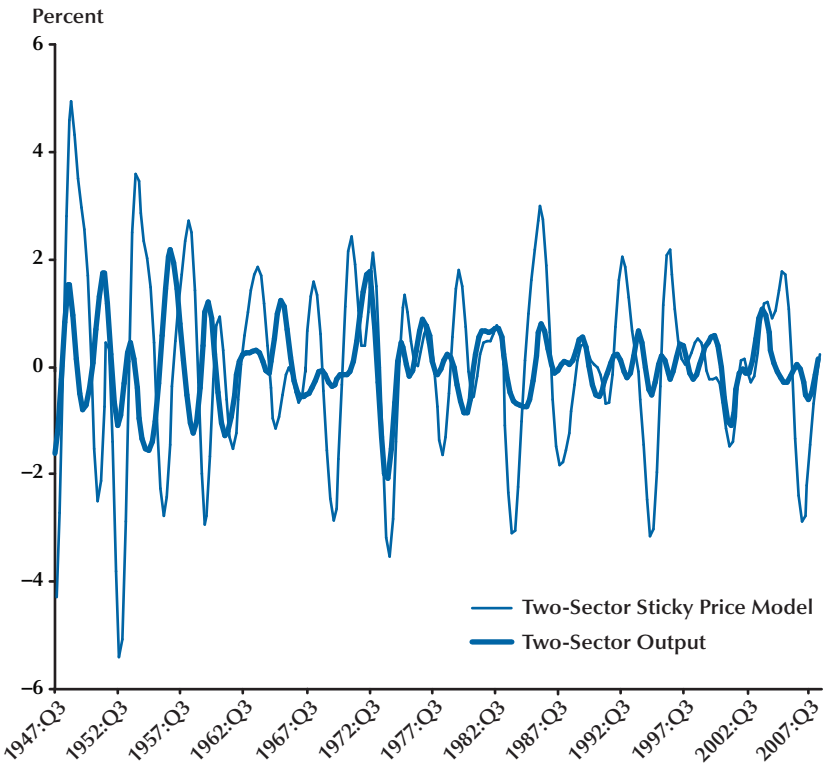
Of course, the welfare benefits of such policies remain controversial. Lucas (1987, 2003) famously argued that, given the fluctuations we observe, the welfare gains from additional stabilization of the economy are likely to be small. In particular, given standard preferences and the observed variance of consumption (around a linear trend), a representative consumer would be willing to reduce his or her average consumption by only about ½ of 1/10th of 1 percent in exchange for eliminating all remaining variability in consumption. Note that this calculation does not necessarily imply that stabilization policy does not matter, because the calculation takes as given the stabilization policies implemented in the past. Stabilization policies might well have been valuable—for example, in eliminating recurrences of the Great Depression or by minimizing the frequency of severe recessions—but additional stabilization might not offer large benefits.

This calculation amounts to some $5 billion per year in the United States, or about $16 per person. Compared with the premiums we pay for very partial insurance (e.g., for collision coverage on our cars), this is almost implausibly low. Any politician would surely vote to pay $5 billion for a policy that would eliminate recessions.

Hence, a sizable literature considers ways to obtain larger costs of business cycle fluctuations, with mixed results. Arguments in favor of stabilization include Galí, Gertler, and López-Salido (2007), who argue that the welfare effects of booms

Figure 3

Output Gap and Sticky Price Output



NOTE: Bandpass-filtered (6 to 32 quarters) output from two-sector sticky price model and the corresponding output gap (defined as sticky price output less flexible price output).

and recessions may be asymmetric. In particular, because of wage and price markups, steady-state employment and output are inefficiently low in their model, so that the costs of fluctuations depend on how far the economy is from full employment. Recessions are particularly costly—welfare falls by more during a business cycle downturn than it rises during a symmetric expansion. Barlevy (2004) argues in an endogenous-growth framework that stabilization might increase the economy's long-run growth rate; this allowed him to obtain very large welfare effects from business cycle volatility.

This discussion of welfare effects highlights that much work remains to understand the desirability of observed fluctuations, the ability of policy to smooth the undesirable fluctuations in the output gap, and the welfare benefits of such policies.

WHAT IS CURRENT POTENTIAL OUTPUT GROWTH?

Consider the current situation, as of late 2008: Is potential output growth relatively high, relatively low, or close to its steady-state value?[24] The answer is important for policymakers, where statements by the Federal Open Market Committee (FOMC) participants have emphasized the impor-

[24] We could, equivalently, discuss the magnitude or even sign of the output gap, which is naturally defined in levels. The level is the integral of the growth rates, of course, and growth rates make it a little easier to focus, at least implicitly, on how the output gap is likely to change over time.

tance of economic weakness in reducing inflationary pressures.[25] Moreover, a discussion of the issue highlights some of what we know, and do not know, about potential output. Some of the considerations are closely linked to earlier points we have made, but these considerations also allow a discussion of other issues that are not included in the simple models discussed here.

Several arguments suggest that potential output growth might currently be running at a relatively rapid pace. First, and perhaps most importantly, TFP growth has been relatively rapid from the end of 2006 through the third quarter of 2008 (see Table 2). During this period output growth itself was relatively weak, and hours per worker were generally falling; hence, following the logic in Basu, Fernald, and Kimball (2006), factor utilization appears to have been falling as well. As a result, in both the consumption and the investment sectors, utilization-adjusted TFP (from Fernald, 2008) has grown at a more rapid pace than its post-1995 average. This fast pace has occurred despite the reallocations of resources away from housing and finance and the high level of financial stress.

Second, substantial declines in wealth are likely to increase desired labor supply. Most obviously, housing wealth has fallen and stock market values have plunged; but tax and expenditure policies aimed at stabilizing the economy could also suggest a higher present value of taxes. Declining wealth has a direct, positive effect on labor supply. In addition, as the logic of Campbell and Hercowitz (2006) would imply, rising financial stress could lead to increases in labor supply as workers need to acquire larger down payments for purchases of consumer durables. And if there is habit persistence in consumption, workers might also seek, at least temporarily, to work more hours to smooth the effects of shocks to gasoline and food prices.

Nevertheless, there are also reasons to be concerned that potential output growth is currently lower than its pace over the past decade or so. First, Phelps (2008) raises the possibility that because of a sectoral shift away from housing-related activities and finance, potential output growth is temporarily low and the natural rate of unemployment is temporarily high. Although qualitatively suggestive, it is unclear that the sectoral shifts argument is quantitatively important. For example, Valletta and Cleary (2008) look at the (weighted) dispersion of employment growth across industries, a measure used by Lilien (1982). They find that as of the third quarter of 2008, "the degree of sectoral reallocation…remains low relative to past economic downturns." Valletta and Cleary (2008) also consider job vacancy data, which Abraham and Katz (1986) suggest could help distinguish between sectoral shifts and pure cyclical increases in unemployment and employment dispersion. The basic logic is that in a sectoral shifts story, expanding firms should have high vacancies that partially or completely offset the low vacancies in contracting firms. Valletta and Cleary find that the vacancy rate has been steadily falling since late 2006.[26]

Third, Bloom (2008) argues that uncertainty shocks are likely to lead to a sharp decline in output. As he puts it, there has been "a huge surge in uncertainty that is generating a rapid slow-down in activity, a collapse of banking preventing many of the few remaining firms and consumers that want to invest from doing so, and a shift in the political landscape locking in the damage through protectionism and anti-competitive policies" (p. 4). His argument is based on the model simulations in Bloom (2007), in which an increase in macro uncertainty causes firms to temporarily pause investment and hiring. In his model, productivity growth also falls temporarily because of reduced reallocation from lower- to higher-productivity establishments.

Fourth, the credit freeze could directly reduce productivity-improving reallocations, along the lines suggested by Bloom (2007), as well as Eisfeldt and Rampini (2006). Eisfeldt and Rampini argue that, empirically, capital reallocation is procycli-

[25] For example, in the minutes from the September 2008 FOMC meeting, participants forecast that over time "increased economic slack would tend to damp inflation" (Board of Governors, 2008).

[26] Valletta and Cleary do find some evidence that the U.S. Beveridge curve might have shifted out in recent quarters relative to its position from 2000 to 2006.

cal, whereas the benefits (reflecting cross-sectional dispersion of marginal products) are countercyclical. These observations suggest that the informational and contractual frictions, including financing constraints, are higher in recessions. The situation as of late 2008 is one in which financing constraints are particularly severe, which is likely to reduce efficient reallocations of both capital and labor.

Fifth, there could be other effects from the seize-up of financial markets in 2008. Financial intermediation is an important intermediate input into production in all sectors. If it is complementary with other inputs (as in Jones, 2008), for example, you need access to the commercial paper market to finance working capital needs—then it could lead to substantial disruptions of real operations.

Finally, the substantial volatility in commodity prices, especially oil, in recent years could affect potential output. That said, although oil is a crucial intermediate input into production, changes in oil prices do not have a clear-cut effect on TFP, measured as domestic value added relative to primary inputs of capital and labor. They might, nevertheless, influence equilibrium output by affecting equilibrium labor supply. Blanchard and Galí (2007) and Bodenstein, Erceg, and Guerrieri (2008), however, are two recent analyses in which, because of (standard) separable preferences, there is no effect on flexible price GDP or employment from changes in oil prices. So there is no a priori reason to expect fluctuations in oil prices to have a substantial effect on the level or growth rate of potential output.

A difficulty for all these arguments that potential output growth might be temporarily low is the observation already made, that productivity growth (especially after adjusting for utilization) has, in fact, been relatively rapid over the past seven quarters.

It is possible the productivity data have been mismeasured in recent quarters.[27] Basu, Fernald, and Shapiro (2001) highlight variations in disruption costs associated with tangible investment. Comparing 2004:Q4–2006:Q4 (when productivity growth was weak) with 2006:Q4–2008:Q3 (when productivity was strong), growth in business fixed investment was very similar, suggesting that time-varying disruption costs probably explain little of the recent variation in productivity growth rates.

Basu et al. (2004) and Oliner, Sichel, and Stiroh (2007) discuss the role of mismeasurement associated with *in*tangible investments, such as organizational changes associated with IT. With greater concerns about credit and cash flow, firms might have deferred organizational investments and reallocations; in the short run, such deferral would imply faster measured productivity growth, even if true productivity growth (in terms of total output, the sum of measured output plus unobserved intangible investment) were constant. Basu et al. (2004) argue for a link between observed investments in computer equipment and unobserved intangible investments in organizational change. Growth in computer and software investment does not show a notable difference between the 2004:Q4–2006:Q4 and 2006:Q4–2008:Q3 periods. If anything, the investment rate was higher in the latter period—so that this proxy again does not imply mismeasurement.

Given wealth effects on labor supply and strong recent productivity performance—along with the failure of typical proxies for mismeasurement to explain the productivity performance—there are reasons for optimism about the short-run pace of potential output growth. Nevertheless, the major effects of the adverse shocks on potential output seem likely to be ahead of us. For example, the widespread seize-up of financial markets has been especially pronounced only in the second half of 2008. We expect that as the effects of the collapse in financial intermediation, the surge in uncertainty, and the resulting declines in factor reallocation play out over the next several years, short-run potential output growth will be constrained relative to where it otherwise would have been.

[27] Note also that the data are all subject to revision. For example, the annual revision in 2009 will revise data from 2006 forward. In addition, labor-productivity data for the nonfinancial corporate sector, which is based on income-side rather than expenditure-side data, show less of a slowdown in 2005 and 2006 and less of a pickup since then. That said, even the nonfinancial corporate productivity numbers have remained relatively strong in the past few years.

CONCLUSION

This article has highlighted a few things we think we know about potential output—namely, the importance in both the short run and the long run of rapid technological change in producing equipment investment goods and the likely time variation in the short-run growth rate of potential. Our discussion of these points has, of course, pointed toward some of the many things we do not know.

Taking a step back, we have advocated thinking about policy in the context of explicit models that suggest ways to think about the world economy, including potential output. But there is an important interplay between theory and measurement, as the discussion suggests. Every day, policymakers grapple with challenges that are not present in the standard models. Not only do they not know the true model of the economy, they also do not know the current state variables or the shocks with any precision; and the environment is potentially nonstationary, with the continuing question of whether structural change (e.g., parameter drift) has occurred. Theory (and practical experience) tells us that our measurements are imperfect, particularly in real time. Not surprisingly, central bankers look at many of the real-time indicators and filter them analytically—relying on theory and experience. Estimating potential output growth is one modest and relatively transparent example of this interplay between theory and measurement.

REFERENCES

Abraham, Katharine G. and Katz, Lawrence K. "Cyclical Unemployment: Sectoral Shifts or Aggregate Disturbances?" *Journal of Political Economy*, June 1986, *94*(3), pp. 507-22.

Alexopoulos, Michelle. "Read All About It! What Happens Following a Technology Shock." Working Paper, University of Toronto, April 2006.

Barlevy, Gadi. "The Cost of Business Cycles Under Endogenous Growth." *American Economic Review*, September 2004, *94*(4), pp. 964-90.

Basu, Susanto and Fernald, John G. "Returns to Scale in U.S. Production: Estimates and Implications." *Journal of Political Economy*, April 1997, *105*(2), pp. 249-83.

Basu, Susanto and Fernald, John G. "Aggregate Productivity and Aggregate Technology." *European Economic Review*, June 2002, *46*(6), pp. 963-91.

Basu, Susanto; Fernald, John G. and Kimball, Miles S. "Are Technology Improvements Contractionary?" *American Economic Review*, December 2006, *96*(5), 1418-48.

Basu, Susanto; Fernald, John G.; Oulton, Nicholas and Srinivasan, Sylaja. "The Case of the Missing Productivity Growth: Or, Does Information Technology Explain Why Productivity Accelerated in the United States but Not the United Kingdom?" in M. Gertler and K. Rogoff, eds., *NBER Macroeconomics Annual 2003*. Cambridge, MA: MIT Press, 2004, pp. 9-63.

Basu, Susanto; Fernald, John G. and Shapiro, Matthew D. "Productivity Growth in the 1990s: Technology, Utilization, or Adjustment?" *Carnegie-Rochester Conference Series on Public Policy*, December 2001, *55*(1), pp. 117-65.

Basu, Susanto; Fisher, Jonas; Fernald, John G. and Miles, Kimball S. "Sector-Specific Technical Change." Unpublished manuscript, University of Michigan, 2008.

Basu, Susanto and Kimball, Miles. "Long Run Labor Supply and the Elasticity of Intertemporal Substitution for Consumption." Unpublished manuscript, University of Michigan, October 2002; www-personal.umich.edu/~mkimball/pdf/cee_oct02-3.pdf.

Beaudry, Paul; Doms, Mark and Lewis, Ethan. "Endogenous Skill Bias in Technology Adoption: City-Level Evidence from the IT Revolution." Working Paper No. 2006-24, Federal Reserve Bank of San Francisco, August 2006; www.frbsf.org/publications/economics/papers/2006/wp06-24bk.pdf.

Blanchard, Olivier, J. and Galí, Jordi. "The Macroeconomic Effects of Oil Price Shocks: Why

Are the 2000s So Different from the 1970s?" Working Paper No. 07-01, MIT Department of Economics, August 18, 2007.

Bloom, Nicholas. "The Impact of Uncertainty Shocks." NBER Working Paper No. 13385, National Bureau of Economic Research, September 2007; www.nber.org/papers/w13385.pdf.

Bloom, Nicholas. "The Credit Crunch May Cause Another Great Depression." Stanford University Department of Economics, October 8, 2008; www.stanford.edu/~nbloom/CreditCrunchII.pdf.

Board of Governors of the Federal Reserve System. Minutes of the Federal Open Market Committee. September 16, 2008; www.federalreserve.gov/monetarypolicy/fomcminutes20080916.htm.

Bodenstein, Martin; Erceg, Christopher E. and Guerrieri, Luca. "Optimal Monetary Policy with Distinct Core and Headline Inflation Rates." International Finance Discussion Papers 941, Board of Governors of the Federal Reserve System, August 2008; www.federalreserve.gov/pubs/ifdp/2008/941/ifdp941.pdf.

Calvo, Guillermo. "Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics*, September 1983, *12*(3), pp. 383-98.

Campbell, Jeffrey and Hercowitz, Zvi. "The Role of Collateralized Household Debt in Macroeconomic Stabilization." Working Paper No. 2004-24, Federal Reserve Bank of Chicago, revised December 2006; www.chicagofed.org/economic_research_and_data/publication_display.cfm?Publication=6&year=2000%20AND%202005.

Card, David. "Intertemporal Labor Supply: An Assessment," in C.A. Sims, ed., *Advances in Econometrics*. Volume 2, Sixth World Congress. New York: Cambridge University Press, 1994, pp. 49-80.

Congressional Budget Office. "CBO's Method for Estimating Potential Output: An Update." August 2001; www.cbo.gov/ftpdocs/30xx/doc3020/PotentialOutput.pdf.

Congressional Budget Office. "A Summary of Alternative Methods for Estimating Potential GDP." March 2004; www.cbo.gov/ftpdocs/51xx/doc5191/03-16-GDP.pdf.

Congressional Budget Office. "Key Assumptions in CBO's Projection of Potential Output" (by calendar year) in The Budget and Economic Outlook: An Update. September 2008, Table 2-2; www.cbo.gov/ftpdocs/97xx/doc9706/Background_Table2-2.xls.

Cummins, Jason G. and Violante, Giovanni L. "Investment-Specific Technical Change in the US (1947-2000): Measurement and Macroeconomic Consequences." *Review of Economic Dynamics*, April 2002, *5*(2), pp. 243-84.

DeLong, J. Bradford. "Productivity Growth in the 2000s," in M. Gertler and K. Rogoff, eds., *NBER Macroeconomics Annual 2002*. Cambridge, MA: MIT Press, 2003.

Edge, Rochelle M.; Kiley, Michael T. and Laforte, Jean-Philippe. "Natural Rate Measures in an Estimated DSGE Model of the U.S. Economy." Finance and Economics Discussion Series 2007-08, Board of Governors of the Federal Reserve System, March 26, 2007; www.federalreserve.gov/pubs/feds/2007/200708/200708pap.pdf.

Eisfeldt, Andrea and Rampini, Adriano. "Capital Reallocation and Liquidity." *Journal of Monetary Economics*, April 2006, *53*(3), pp. 369-99.

Elsby, Michael and Shapiro, Matthew. "Stepping Off the Wage Escalator: A Theory of the Equilibrium Employment Rate." Unpublished manuscript, April 2008; www.eief.it/it/files/2008/04/stepping-off-2008-04-01.pdf.

Fernald, John G. "A Quarterly Utilization-Adjusted Measure of Total Factor Productivity." Unpublished manuscript, 2008.

Fisher, Jonas. "The Dynamic Effects of Neutral and Investment-Specific Technology Shocks." *Journal of Political Economy*, June 2006, *114*(3), pp. 413-52.

Galí, Jordi; Gertler, Mark and Lopez-Salido, David J. "Markups, Gaps, and the Welfare Costs of Business

Fluctuations." *Review of Economics and Statistics*, November 2007, *89*, pp. 44-59.

Gordon, Robert J. *The Measurement of Durable Goods Prices*. Chicago: University of Chicago Press, 1990.

Gordon, Robert J. "Does the 'New Economy' Measure up to the Great Inventions of the Past?" *Journal of Economic Perspectives*, Fall 2000, *4*(14), pp. 49-74.

Gordon, Robert J. "Future U.S. Productivity Growth: Looking Ahead by Looking Back." Presented at the Workshop at the Occasion of Angus Maddison's 80th Birthday, *World Economic Performance: Past, Present, and Future*, University of Groningen, Netherlands, October 27, 2006.

Greenwood, Jeremy; Hercowitz, Zvi and Krusell, Per. "Long-Run Implications of Investment-Specific Technological Change." *American Economic Review*, June 1997, *87*(3), pp. 342-62.

Hansen, Gary. "Indivisible Labor and the Business Cycle." *Journal of Monetary Economics*, November 1985, *16*, pp. 309-37.

Jones, Chad. "R&D-Based Models of Economic Growth." *Journal of Political Economy*, August 1995, *103*, pp. 759-84.

Jones, Chad. "Sources of U.S. Economic Growth in a World of Ideas." *American Economic Review*, March 2002, *92*(1), pp. 220-39.

Jones, Chad. "Intermediate Goods and Weak Links: A Theory of Economic Development." NBER Working Paper No. 13834, National Bureau of Economic Research, September 2008; www.nber.org/papers/w13834.pdf.

Jorgenson, Dale W.; Gollop, Frank M. and Fraumeni, Barbara M. *Productivity and U.S. Economic Growth*. Cambridge, MA: Harvard University Press, 1987.

Jorgenson, Dale W. "Information Technology and the U.S. Economy." *American Economic Review*, March 2001, *91*(1), pp. 1-32.

Jorgenson, Dale W.; Ho, Mun S. and Stiroh, Kevin J. "A Retrospective Look at the U.S. Productivity Growth Resurgence." *Journal of Economic Perspectives*, Winter 2008, *22*(1), pp. 3-24.

Justiniano, Alehandro and Primiceri, Giorgio. "Potential and Natural Output." Unpublished manuscript, Northwestern University, June 2008; http://faculty.wcas.northwestern.edu/~gep575/JPgap8_gt.pdf.

King, Robert G.; Plosser, Charles I. and Rebelo, Sergio T. "Production, Growth and Business Cycles: I. The Basic Neoclassical Model." *Journal of Monetary Economics*, 1988, *21*(2-3), pp. 195-232.

Kuttner, Kenneth. "Estimating Potential Output as a Latent Variable." *Journal of Business and Economic Statistics*, July 1994, *12*(3), pp. 361-68.

Kydland, Finn E. and Prescott, Edward C. "Rules Rather than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy*, June 1977, *85*(3), pp. 473-92.

Laubach, Thomas and Williams, John C. "Measuring the Natural Rate of Interest." *Review of Economics and Statistics*, November 2003, *85*(4), pp. 1063-70.

Lilien, David M. "Sectoral Shifts and Cyclical Unemployment." *Journal of Political Economy*, August 1982, *90*(4), pp. 777-93.

Lucas, Robert E. Jr. "Econometric Policy Evaluation: A Critique." *Carnegie-Rochester Conference Series on Public Policy*, 1976, *1*(1), pp. 19-46.

Lucas, Robert E. Jr. *Models of Business Cycles*. Oxford: Basil Blackwell Ltd, 1987.

Lucas, Robert E. Jr. "Macroeconomic Priorities." *American Economic Review*, March 2003, *93*(1), pp. 1-14.

Mulligan, Casey. "Aggregate Implications of Indivisible Labor." *Advances in Macroeconomics*, 2001, *1*(1), Article 4; www.bepress.com/cgi/viewcontent.cgi?article=1007&context=bejm.

Neiss, Katherine and Nelson, Edward. "Inflation Dynamics, Marginal Cost, and the Output Gap: Evidence from Three Countries." *Journal of Money,*

Credit, and Banking, December 2005, *37*(6), pp. 1019-45.

Okun, Arthur, M. *The Political Economy of Prosperity*. Washington, DC: Brookings Institution, 1970.

Oliner, Stephen D. and Sichel, Daniel E. "The Resurgence of Growth in the Late 1990s: Is Information Technology the Story?" *Journal of Economic Perspectives*, Fall 2000, *14*(4), pp. 3-22.

Oliner, Stephen D. and Sichel, Daniel E. "Information Technology and Productivity: Where Are We Now and Where Are We Going?" Federal Reserve Bank of Atlanta *Economic Review*, Third Quarter 2002, pp. 15-44; www.frbatlanta.org/filelegacydocs/oliner_sichel_q302.pdf.

Oliner, Stephen D.; Sichel, Daniel and Stiroh, Kevin. "Explaining a Productive Decade." *Brookings Papers on Economic Activity*, 2007, *1*, pp. 81-137.

Organisation of Economic Co-operation and Development. Revisions of Quarterly Output Gap Estimates for 15 OECD Member Countries. September 26, 2008; www.oecd.org/dataoecd/15/6/41149504.pdf.

Phelps, Edmund S. "U.S. Monetary Policy and the Prospective Structural Slump." Presented at the 7th Annual BIS Monetary Policy Conference, Lucerne, June 26-27, 2008; www.bis.org/events/conf080626/phelps.pdf.

Rogerson, Richard. "Indivisible Labor, Lotteries and Equilibrium." *Journal of Monetary Economics*, January 1988, *21*(1), pp. 3-16.

Rotemberg, Julio J. "Stochastic Technical Progress, Nearly Smooth Trends and Distinct Business Cycles." NBER Working Paper 8919, National Bureau of Economic Research, May 2002; papers.ssrn.com/sol3/papers.cfm?abstract_id=310466.

Smets, Frank and Wouters, Rafael. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review*, June 2007, *97*(3), pp. 586-606.

Survey of Professional Forecasters. Survey from First Quarter 2008. February 12, 2008; www.philadelphiafed.org/ research-and-data/real-time-center/survey-of-professional-forecasters/2008/spfq108.pdf.

Solow, Robert M. "Is There a Core of Usable Macroeconomics We Should All Believe In?" *American Economic Review*, May 1997, *87*(2), pp. 230-32.

Valletta, Robert and Cleary, Aisling. "Sectoral Reallocation and Unemployment." Federal Reserve Bank of San Francisco *FRBSF Economic Letter*, No. 2008-32, October 17, 2008; www.frbsf.org/publications/economics/letter/2008/el2008-32.pdf.

Whelan, Karl. "A Two-Sector Approach to Modeling U.S. NIPA Data." *Journal of Money, Credit, and Banking*, August 2003, *35*(4), pp. 627-56.

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

APPENDIX

A SIMPLE TWO-SECTOR STICKY PRICE MODEL[28]

Households

The economy is populated by a representative household which maximizes its lifetime utility, denoted as

$$\max E_0 \sum_{t=0}^{\infty} u(C_t, L_t),$$

where C_t is consumption of a constant elasticity of substitution basket of differentiated varieties

$$C_t = \left[\int_0^1 C(z)^{\frac{\xi-1}{\xi}} dz \right]^{\frac{\xi}{\xi-1}}$$

and L_t is labor effort. u, the period felicity function, takes the following form:

$$u_t = \ln C_t - \frac{L_t^{\eta+1}}{\eta+1},$$

where η is the inverse of the Frisch elasticity of labor supply. The maximization problem is subject to several constraints. The flow budget constraint, in nominal terms, is the following:

$$B_t + P_t^I I_t + P_t^C C_t = W_t L_t + R_t K_{t-1} + (1 + i_{t-1}) B_{t-1} + \Delta,$$

where

$$I = \left[\int_0^1 I(z)^{\frac{\xi-1}{\xi}} dz \right]^{\frac{\xi}{\xi-1}}.$$

The price indices are defined as follows:

$$P_t^C = \left(\int_0^1 P_t^C(z)^{1-\xi} dz \right)^{\frac{1}{1-\xi}}$$

$$P_t^I = \left(\int_0^1 P_t^I(z)^{1-\xi} dz \right)^{\frac{1}{1-\xi}}.$$

Moreover,

(A1) $$K_t = I_t + (1-\delta) K_{t-1}$$

(A2) $$L_t = L_t^C + L_t^I$$

(A3) $$K_{t-1} = K_t^C + K_t^I.$$

[28] The appendix was written primarily by Alessandro Barattieri.

Notice that total capital is predetermined, while sector-specific capital is free to move in each period. To solve the problem, we write the Lagrangian as

$$L = \ln C_t - \frac{L_t^{\eta+1}}{\eta+1} + \ldots - \Lambda_t\left(B_t + P_t^I I_t + P_t^C C_t - W_t L_t - R_t K_{t-1} - \left(1+i_{t-1}\right)B_{t-1} - \Delta_t\right)$$
$$-\beta E_t \Lambda_{t+1}\left(B_{t+1} + P_{t+1}^I I_{t+1} + P_{t+1}^C C_{t+1} - W_{t+1}L_{t+1} - R_{t+1}K_t - \left(1+i_t\right)B_t - \Delta_{t+1}\right) - \ldots$$

The first-order conditions of the maximization problem for consumption, nominal bond, labor, and capital are as follows:

(A4) $$\frac{1}{C_t} = P_t^C \Lambda_t$$

(A5) $$\Lambda_t = \beta E_t\left[\left(1+i_t\right)\Lambda_{t+1}\right]$$

(A6) $$L^\eta = \Lambda_t w_t$$

(A7) $$\Lambda_t P_t^I = \beta E_t\left[\Lambda_{t+1}\left(R_{t+1} + P_{t+1}^I\left(1-\delta\right)\right)\right].$$

Table A1 provides baseline calibrations for all parameters.

Table A1

Baseline Calibration

Parameter	Value	Parameter	Value
β	0.99	*INV_SHARE*	0.2
η	0.25	*C_SHARE*	0.8
α_C	0.3	L^I/L	0.2
α_I	0.3	L^C/L	0.8
δ	0.025	K^I/K	0.2
Γ^C	1.1	K^C/K	0.8
Γ^I	1.1	ρ_i	0.8
θ_C	0.75	ϕ_π	1.5
θ_I	0.75	ϕ_μ	0.5
ζ_C	$\frac{\left(1-\theta_C\right)\left(1-\beta\theta_C\right)}{\theta_C}$	ρ_C	0.99
		ρ_I	0.99
		$\sigma_{\varepsilon_t^C}$	1
ζ_I	$\frac{\left(1-\theta_I\right)\left(1-\beta\theta_I\right)}{\theta_I}$	$\sigma_{\varepsilon_t^I}$	1
		σ_{v_t}	1

Firms

Both sectors are characterized by a unitary mass of atomistic monopolistically competitive firms. Production functions are Cobb-Douglas (possibly with different factor intensities). Productivity in the two sectors is represented by two AR(1) processes. The cost minimization problem for the firms z operating in the consumption and investment sectors can be expressed, in nominal terms, as

$$\text{Min } W_t L_t^C(z) + R_t K_t^C(z)$$
$$\text{s.t. } Y_t^C(z) = A_t^C \left(K_t^C(z)\right)^{\alpha^C} \left(L_t^C(z)\right)^{1-\alpha^C} - \Phi^C$$

and analogously as

$$\text{Min } W_t L_t^I(z) + R_t K_t^I(z)$$
$$\text{s.t. } I_t = A_t^I \left(K_t^I(z)\right)^{\alpha^I} \left(L_t^I(z)\right)^{1-\alpha^I} - \Phi^I.$$

Calling μ^i with $i = C, I$ the multiplier attached to the minimization problem, reflecting nominal marginal cost, we can express the factor demand as follows, where we omit z assuming a symmetric equilibrium:

$$\frac{W_t}{\mu_t^C} = (1-\alpha) A_t^C \left(K_t^C\right)^{\alpha^C} \left(L_t^C\right)^{-\alpha^C}$$
$$\frac{R_t}{\mu_t^C} = \alpha A_t^C \left(K_t^C\right)^{\alpha^C - 1} \left(L_t^C\right)^{1-\alpha^C}$$
$$\frac{W_t}{\mu_t^I} = (1-\alpha) A_t^I \left(K_t^I\right)^{\alpha^I} \left(L_t^I\right)^{-\alpha^I}$$
$$\frac{R_t}{\mu_t^I} = \alpha A_t^I \left(K_t^I\right)^{\alpha^I - 1} \left(L_t^I\right)^{1-\alpha^I}.$$

Taking the ratio for each sector, we get

(A8) $$\frac{K_t^C}{L_t^C} = \frac{\alpha}{1-\alpha} \frac{W_t}{R_t}$$

(A9) $$\frac{K_t^I}{L_t^I} = \frac{\alpha}{1-\alpha} \frac{W_t}{R_t}.$$

Inflation rates are naturally defined as

(A10) $$\Pi_t^j = \frac{P_t^j}{P_{t-1}^j}.$$

Finally, given the Cobb-Douglas assumption, it is possible to express the nominal marginal cost as follows:

(A11) $$MC^j = \frac{1}{A^j} \frac{1}{f(\alpha)} R^\alpha W^{1-\alpha} \left(Y^j + \Phi^j\right),$$

with $j = C, I$.

We introduce nominal rigidities through standard Calvo (1983) pricing. Instead of writing the rather complex equations for the price levels in the C and I sectors, we jump directly to the log-linearized Calvo equations for the evolution of inflation rates, equations (25) and (26) below.

Monetary Policy

Monetary policy is conducted through a Taylor-type rule with a smoothing parameter and reaction to inflation and marginal cost. Again, we write the Taylor rule directly in log-linearized form, as equation (A32) below.

Equilibrium

Beyond the factor market–clearing conditions already expressed, equilibrium also requires a bond market–clearing condition ($B = 0$), a consumption goods market–clearing condition ($Y^C = C$), and an aggregate adding-up condition ($C + I = Y$). (By Walras's law, we drop the investment market–clearing condition.)

The Linearized Model

The equations of the model linearized around its nonstochastic steady state are represented by equations (A12) through (A36), which are 25 equations for the 25 unknown endogenous variables, c, l^I, l^C, l, k^C, k^I, k, λ, w, w^C, r, i, y^C, I, y, p^I, p^C, π, π^C, π^I, μ, μ^I, μ^C, a^C, a^I, as follows:

(A12) $$k_t = \delta I_t + (1-\delta) k_{t-1}$$

(A13) $$\frac{L^I}{L} l_t^I + \frac{L^C}{L} l_t^C = l_t$$

(A14) $$\frac{K^I}{K} k_t^I + \frac{K^C}{K} k_t^C = k_{t-1}$$

(A15) $$-c_t = \lambda_t + p_t^C$$

(A16) $$\lambda_t = i_t + \lambda_{t+1}$$

(A17) $$\eta l = \lambda_t + w_t$$

(A18) $$\lambda_t + p_t^I = \lambda_{t+1} + \left[1 - \beta(1-\delta)\right] r_{t+1} + \beta(1-\delta) p_{t+1}^I$$

(A19) $$y_t^C = \Gamma^C \left(a_t^C + \alpha^C k_t^C + \left(1-\alpha^C\right) l_t^C\right)$$

(A20) $$I_t = \Gamma^I \left(a_t^I + \alpha^I k_t^I + \left(1-\alpha^I\right) l_t^I\right)$$

(A21) $$k_t^C + r_t = w_t + l_t^C$$

(A22) $$k_t^I + r_t = w_t + l_t^I$$

(A23) $$\mu_t^C = \alpha r_t + (1-\alpha) w_t - a_t^C$$

(A24) $$\mu_t^I = \alpha r_t + (1-\alpha) w_t - a_t^I$$

(A25) $$\pi_t^C = \beta \pi_{t+1}^C + \zeta\left(\mu_t^C - p^C\right)$$

(A26) $$\pi_t^I = \beta \pi_{t+1}^I + \zeta\left(\mu_t^I - p^I\right)$$

(A27) $$\pi_t^C = p_t^C - p_{t-1}^C$$

(A28) $$\pi_t^I = p_t^C - p_{t-1}^I$$

(A29) $$\mu_t = C_share \cdot \mu_t^C + INV_share \cdot \mu_t^I$$

(A30) $$\pi_t = C_share \cdot \pi_t^C + INV_share \cdot \pi_t^I$$

(A31) $$w_t^C = w_t - p_t^C$$

(A32) $$i_t = \rho_i i_{t-1} + (1-\rho_i)\left(\phi_\pi \pi_t + \phi_\mu \mu_t\right)$$

(A33) $$y_t = C_share \cdot c_t + INV_share \cdot I_t$$

(A34) $$y_t^C = c_t$$

(A35) $$a_t^C = \rho_C a_{t-1}^C + \varepsilon_t^C$$

(A36) $$a_t^I = \rho_I a_{t-1}^I + \varepsilon_t^I \ .$$

Commentary

Rodolfo E. Manuelli

Basu and Fernald (2009) describe and evaluate alternative theoretical models of potential output to provide a frame of reference for policy analysis. They also discuss what is (and what is not) known about potential output and illustrate their approach by estimating a two-sector model with price rigidities.

I find the overall theme—that models ought to be used to guide policy choices—important and a welcome reminder of the value of using a consistent framework for policy evaluation. I wholeheartedly agree with the approach. When it comes to specifics, they conclude that to capture some essential features of the U.S. economy, the standard one-sector model should be abandoned in favor of a two-sector model with differential technological change. Here, I am not totally convinced by their arguments. The second major point that they argue—and I fully agree with them here—is that any useful notion of potential output cannot be assumed to be properly described by a smooth trend, and it is likely to fluctuate even in the short run. As before, their choice of model and the empirical strategy they use are subject to debate.

THE LONG RUN: WHAT SIMPLE MODEL MATCHES THE DATA?

Basu and Fernald argue that the appropriate notion of potential output is the steady state of an economy with no distortions. They consider two models: a standard one-sector model and a two-sector model with differential technological change across sectors. They derive the steady-state predictions in each case and confront the theoretical predictions about capital deepening—defined as the contribution of the increase in capital per worker to output—with the data. They conclude that the two-sector model, which allows for a change in the price of capital, outperforms the simple one-sector model.

At this level of abstraction, it is not easy to pick a winner. Basu and Fernald base their preference for the two-sector model on two different arguments. First, they show that in the data the relative price of capital has decreased substantially, which is inconsistent with the one-sector model. Second, they highlight the ability of the two-sector model to account for the low contribution of capital in the period of productivity slowdown.

Basu and Fernald's first argument—the change in the price of capital—is not completely persuasive. There is no discussion that capital has become cheaper, but this does not automatically imply that this fact is of crucial importance. Of necessity, models are abstractions of reality and, by their very nature, will miss some dimension of the data. To be precise, models that account for everything are so complex that they cannot be useful. Thus, adding a sector—which can only improve the ability of the model—cannot determine a winner. It is easy enough to find other

Rodolfo E. Manuelli is a professor of economics at Washington University in St. Louis and a research fellow at the Federal Reserve Bank of St. Louis. The author thanks Yongs Shin for useful conversations on this topic.

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 215-19.
© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

Table 1

Capital's Contribution: Model Prediction/Data

Period	One-sector model	Two-sector model
1948-73	1.2	1.05
1973-95	0.4	1.06
1995-2000	0.6	1.75
2000-07	0.9	1.54
1948-2007	0.9	1.18

changes in relative prices (e.g., some professional services) that would necessitate a third sector to accommodate them, and this approach would logically lead to a complex and useless model.

Basu and Fernald's primary reason for choosing a two-sector model rests in its ability to explain capital deepening. Table 1 presents the two models' predictions for capital's contribution to growth relative to the data for various time periods. Considering the longest available horizon (1948-2007), it is difficult to choose a winner. The one-sector model underpredicts the contribution of capital by 15 percent, while the two-sector model overpredicts it by 18 percent. Depending on the period, one model clearly dominates the other, but I see no reason to emphasize the 1973-95 period (where the two-sector model is a clear winner) over the 2000-07 period (in which the one-sector model dominates).

Basu and Fernald's preferred model is a two-sector version of the Solow growth model. Using data on the relative price of capital, they estimate the productivity growth rates in the general goods and investment goods sectors. Their estimate hinges on the assumption that the technologies in these two sectors are similar. In particular, letting $\alpha_c = \alpha_i$ be the growth rate of total factor productivity (TFP) in sector j, the specification implies that

$$\hat{P}_i - \hat{P}_c = \hat{z}_c - \hat{z}_i.$$

In a version of the model in which the capital shares are allowed to differ across sectors, the relative price of consumption satisfies

Table 2

Price of Capital (1948 = 1)

Period	Model	Data
1973	1.05	1.09
1995	0.90	0.91
2000	0.82	0.87
2007	0.77	0.86

$$\hat{P}_i - \hat{P}_c = \hat{z}_c - \frac{1-\alpha_c}{1-\alpha_i}\hat{z}_i.$$

Valentinyi and Herrendorf (2008) estimate $\alpha_c = \alpha_i$ and $\alpha_i = 0.28$, which implies that, relative to Basu and Fernald's estimate, the productivity growth rate of the investment sector was about 9 percent higher. This implies that their estimates of the contribution of capital deepening must be increased by almost 10 percent, which exaggerates even more the overprediction of the two-sector model relative to the data in the recent past.

Even accepting as a reasonable approximation that $\alpha_c = \alpha_i$, there are two measures of the relative price of investment goods ($p_i = P_i/P_c$) that, according to the model, should coincide. One is given by

$$p_i = M_k\left(\frac{Y/L}{K/L}\right),$$

where $y = Y/K (k = K/L)$ is output (capital) per hour and M_k is a constant under the balanced growth assumption. Thus, the growth rate of the relative price of capital is

$$\hat{p}_i = \hat{y} - \hat{k}. \tag{1}$$

As above, the model implies that

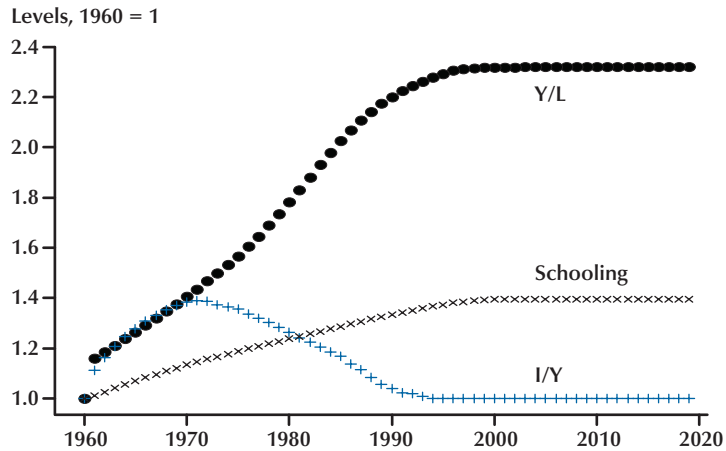
$$\hat{p}_i = \hat{z}_c - \hat{z}_i. \tag{2}$$

The estimates based on equation (1)—using Bureau of Labor Statistics data on output per hour and capital per hour—are presented in the column labeled "Data" in Table 2, while the values from equation (2)—based on model-produced estimates of productivity growth—are labeled "Model."

Because the model-based measure predicts a higher decrease in the price of capital, it is not

Figure 1

Transitional Dynamics: TFP Shock



surprising that the theoretical model tends to overpredict the contribution of capital to output. At this level of abstraction, it is not possible to identify the source of the problem. However, if the effective cost of capital is changing—a violation of the balanced growth assumption—then the "Data" estimate is biased. In any case, the difference should make us cautious about the appropriateness of the model.

Is it clear that balanced growth is a reasonable approximation in the long run, given the length of the horizon covered in the article? It is consistent with the findings of King and Rebelo (1993), who showed that for reasonable parameterizations, the standard growth model converges rather rapidly to its balanced growth path. However, recent work that retains the dynastic specification of preferences but specifies that individual human capital completely depreciates at the time of death (see Manuelli and Seshadri, 2008) has shown that even one-sector models can display very long transitions. Figure 1 presents the impact of a once-and-for-all permanent increase in the level of productivity. From the point of view of this discussion, the interesting result is how long it takes for the model to reach steady state: approximately 30 years. Thus, if human capital that "disappears" when an individual dies (even though dynasties have infinite horizons) is a realistic feature to incorporate in a model, the balanced growth assumption is difficult to justify unless the horizon is very long.

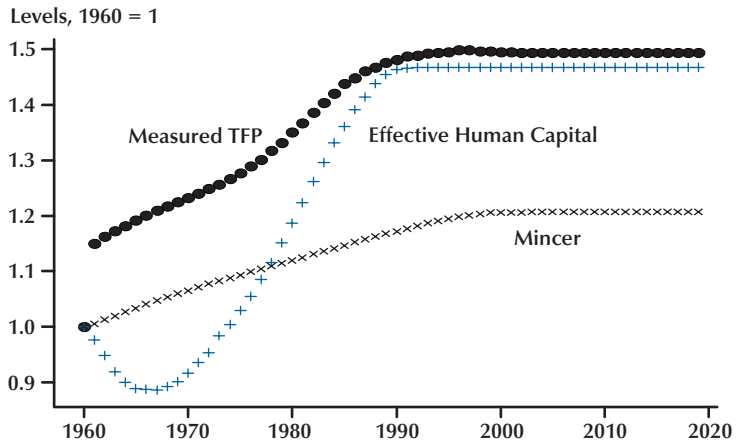
In this case, a second difficulty is associated with the measurement of productivity. In the model analyzed by Manuelli and Seshadri (2008), conventionally measured TFP and actual TFP do not coincide. The divergence is due to the endogenous response of the quality and availability of human capital after a shock. Figure 2 displays measured TFP (computed using the human-capital series labeled "Mincer"), which shows an upward trend—that is, one displaying growth—while "true" TFP jumps in the first period (labeled 1960 in the figure) and remains constant.

In this example, the series labeled "Effective Human Capital" moves in response to a productivity shock. Because measured TFP is simply $zq^{1-\alpha}$, where q is the ratio of Mincer and Effective Human Capital, it follows that measured TFP has a large endogenous component.

Basu and Fernald discuss a variety of scenarios about future productivity growth and trace the implications for output growth. The previous argument suggests that even simple shocks might have a large impact on conventionally measured TFP, which would not be captured in their calcu-

Figure 2

TFP Shock, Effective Human Capital, Mincerian Human Capital, and Measured TFP



lations. Moreover, given the model that they use—essentially one in which the only key decision, saving, is taken as exogenous—any reduced-form representation of the economic variables of interest is an appropriate model to forecast the future, with significantly less structure.

SHORT-RUN CONSIDERATIONS

In this section of their article, Basu and Fernald describe their estimates of technology shocks (i.e., TFP) in a two-sector model and define potential output as the output that would be obtained in the absence of frictions (e.g., price stickiness). Their major finding is that the variability of productivity shocks is high, even at the business cycle frequency, and hence that the prescription that in the short run government policy should try to stabilize output is suspect.

The key question is whether the technology shocks they identify are indeed "purified" of policy-induced fluctuations. I am not totally convinced that simple econometric procedures can effectively isolate TFP shocks, especially given the authors' strong assumption about orthogonality between measured TFP and policy shocks. In particular, it is relatively easy to introduce policies in the Manuelli and Seshadri (2008) model that endogenously change the rate of utilization of human capital (with no change in measured employment) that would appear as changes in technology. Whether these sources of misspecification are important is a question that is difficult to answer using Basu and Fernald's partial-specification approach. As they are aware, some sources of bias can be detected only when they are fully specified in the model.

CONCLUSION

In this discussion, I have taken issue with some of the specific choices made by Basu and Fernald and with their interpretation of the results. I would like to end on a more important note: This paper points policy-based economic research in the right direction because it emphasizes the necessity of being explicit about the assumptions underlying our models. Moreover, by making explicit the economies that are modeled, it is possible to subject the models to a variety of tests. On the other hand, reduced-form atheoretical approaches to policymaking must rely on (often implicit) assumptions to justify their recommendations, and intelligent evaluation of the results is often very difficult, if not outright impossible.

REFERENCES

Basu, Susanto and Fernald, John G. "What Do We Know (And Not Know) About Potential Output?" Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 187-213.

King, Robert G. and Rebelo, Sergio T. "Transitional Dynamics and Economic Growth in the Neoclassical Model." *American Economic Review*, September 1993, *83*(4), pp. 908-31.

Manuelli, Rodolfo E. and Seshadri, Ananth. "Neoclassical Miracles." Working paper, University of Wisconsin–Madison, November 2008; www.econ.wisc.edu/~aseshadr/working_pdf/miracles.pdf.

Valentinyi, Ákos and Herrendorf, Berthold. "Measuring Factor Income Shares at the Sectoral Level." *Review of Economic Dynamics*, October 2008, *11*(4), pp. 820-35.

Issues on Potential Growth Measurement and Comparison: How *Structural* Is the Production Function Approach?

Christophe Cahn and Arthur Saint-Guilhem

This article aims to better understand the factors driving fluctuations in potential output measured by the production function approach (PFA.) To do so, the authors integrate a production function definition of potential output into a large-scale dynamic stochastic general equilibrium (DSGE) model in a fully consistent manner and give two estimated versions based on U.S. and euro-area data. The main contribution of this article is to provide a quantitative and comparative assessment of two approaches to potential output measurement, namely DSGE and PFA, in an integrated framework. The authors find that medium-term fluctuations in potential output measured by the PFA are likely to result from a large variety of shocks, real or nominal. These results suggest that international comparisons of potential growth using the PFA could lead to overstating the role of structural factors in explaining cross-country differences in potential output, while neglecting the fact that different economies are exposed to different shocks over time. (JEL C51, E32, O11, O47)

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 221-40.

International comparisons of potential output growth have received renewed interest in recent years. Lower economic performance in Europe compared with the United States over the past 15 years has generated several publications whose aim is to explain the sources of divergence in economic performance and which question how to enhance economic growth in Europe. In line with the recommendations of the Lisbon strategy, one general conclusion is that *structural* reforms should help to sustain more vigorous growth in Europe and enable European economies to catch up to the United States. Such reforms include labor and product market liberalization, public policies to encourage innovation, and so forth. Examples can be found in most recent International Monetary Fund (IMF) or Organisation for Economic Co-operation and Development (OECD) country reports on European economies. For instance, the 2007 IMF Article IV Staff Report for France (IMF, 2007) typically incorporates, among others, the important conclusion that "economic policy needs to address the root cause of France's growth deficit: the weakness of its supply potential." Against this background, it is important to have a clear view on how potential output is measured and what interpretation can be made of cross-country differences in potential output growth.

Among the different methods of measurement of potential output, the production function approach (PFA) is probably the most widely used. With this approach, output growth is expressed as a sum of the growth of factor inputs (i.e., capital

Christophe Cahn is a doctoral candidate at the Paris School of Economics and an economist with the Banque de France. Arthur Saint-Guilhem is an economist with the European Central Bank. The authors thank Jon Faust for his helpful comments, as well as Richard Anderson and all the participants at the conference.

© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, the regional Federal Reserve Banks, the European Central Bank, the Banque de France, or the Paris School of Economics. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

services and labor input) and a residual (i.e., total factor productivity [TFP] growth). Additional assumptions are made on the potential level of the factors of production. For instance, potential labor input would be calculated by smoothing some variables (such as total population and the participation rate) and by approximating the medium-term equilibrium unemployment rate with the non-accelerating inflation rate of unemployment. The major advantage of the PFA, compared with statistical aggregate methods, is that it provides an economic interpretation of the different factors that drive growth in potential output. This is especially useful in the context of international comparisons. Moreover, conducting additional econometric analysis allows use of the PFA as a framework to capture the impact on potential growth of major changes, such as the pickup in productivity growth that started in the second half of the 1990s in the United States.

However, this approach raises some difficulties. Estimates of the components are bounded by a large degree of uncertainty because analysis results are highly dependent on the choice of modeling of the different components—for instance, how trend growth of TFP is estimated. Another difficulty derives from possible misleading interpretations of potential output as measured by the PFA. First, in the context of international comparisons, cross-country differences in PFA potential output are often given a structural interpretation—say, as being caused by different degrees of rigidities in the labor or good markets, whereas these differences in potential output measures could reflect only the lasting effects of temporary shocks to the economy. This issue is of particular importance because it casts doubt on the ability of the PFA to give a satisfactory picture of the structural components of economic growth. Second, the PFA leaves unidentified the various shocks (supply, demand, monetary shocks, and so on) that are likely to affect potential output in the medium term. This raises some concern about the measurement of output gaps. Indeed, it is not entirely certain that fluctuations in the output gap measured by the PFA reflect only inflation-related shocks. Therefore, the PFA might lead to biased output gap measures that could make them unreliable for the assessment of monetary policy conditions.

An alternative approach to the definition and measurement of potential output can be found in New Keynesian dynamic stochastic general equilibrium (DSGE) models. The recent literature on DSGE models has shown significant progress in developing models that can be applied to the data. Indeed, recent research has shown that estimated DSGE models are able to match the data for key macroeconomic variables and reduced-form vector autoregressions (Smets and Wouters, 2007). In these models, "potential output" is generally defined as the level of output that would prevail in an economy with fully flexible prices and wages. According to the DSGE definition, potential output is therefore the level of output at which prices tend to stabilize. However, the properties of potential output and output gap fluctuations derived from DSGE models can be quite different from the ones derived from the PFA (e.g., Neiss and Nelson, 2005; and Edge, Kiley, and Laforte, 2007). For example, the DSGE measure of potential output can undergo relatively larger fluctuations than potential output derived from the PFA. Similarly, the output gap in DSGE models tends to be less variable than with the PFA measures. One caveat of these papers, however, is that they compare ad hoc PFA measures of potential output with DSGE measures—comparisons that would be enhanced if the PFA measure of potential output were consistent with the model. In this respect, one of the main contributions of our paper is to incorporate the PFA measure of potential output into a DSGE framework in a fully consistent manner. As shown later, adopting such a method reveals that different types of shocks are likely to cause potential output measured by the PFA to fluctuate.

Our goals are twofold: (i) better understanding of the factors driving medium-term fluctuations in the PFA potential output and (ii) providing a quantitative comparison of the PFA versus DSGE measure of potential output. To do so, we build a large-scale DSGE model, calibrate two versions of the model using U.S. and euro-area data, and then integrate into this framework a PFA definition of potential output that is fully consistent with the model. Our PFA is based on previous

work (Cahn and Saint-Guilhem, 2009), where output of the economy is described as a Cobb-Douglas function. In this respect, the main contribution of this paper is to provide a quantitative comparison of these two measures of potential output—the PFA versus the DSGE—in a fully integrated conceptual framework—namely, an economy modeled as a large-scale DSGE model with structural parameters calibrated on U.S. and euro-area data and with an alternative PFA measure of potential output.

A second contribution of this paper is to assess the validity of the structural interpretation of cross-country comparisons of potential output measures given by the PFA. In general, as described previously, potential output estimates based on the PFA suggest significant differences across countries with regard to the sources of potential growth. However, whether these differences can be attributed to structural factors, such as differences in labor market or product market institutions, remains uncertain. Nothing in the PFA guarantees that this is the case. Our present DSGE framework enables us to tackle the issue, given that in such a framework structural differences across two economies translate into differences of magnitude across the various parameters of the model. We can therefore quantify the role of shocks versus the role of structural factors in explaining cross-country differences in potential output measured by the PFA by simulating various counterfactual scenarios for the two model economies.

Our main results first confirm that the PFA and the DSGE definitions of potential output are two different concepts. We find that in an economy modeled with a DSGE framework, medium-term fluctuations in potential output measured by the PFA result from a variety of shocks, such as productivity or monetary shocks. We also find that differences in potential output between two such model economies as measured by the PFA can be attributed not only to structural parameters of the model but also to the role of some transitory shocks, real or nominal, affecting the economies. If we transpose these results into the empirical field, we see two results: (i) PFA measures of potential output also reflect the historical pattern of shocks that affect a given economy, and (ii) international comparisons of potential output using the PFA could lead to overestimating the role of structural factors in explaining cross-country differences in potential output, while neglecting the role of "luck," namely, the fact that different economies are exposed to different histories of stochastic events on which structural policies could not act.

The remainder of this paper is organized as follows. In the next section, we sketch the theoretical specification of our DSGE model. The following section describes how we incorporate and implement into this framework the PFA measure of potential output. We then present and discuss the results of the simulations performed with regard to the decomposition of potential output dynamics into the contributions of the various shocks included in the model. Our summary and conclusion then follow.

A BENCHMARK DSGE MODEL FOR THE UNITED STATES AND THE EURO AREA

In this section, we provide details on the main optimizing behaviors of economic agents—households, firms, and the fiscal and monetary authorities—that lead to building the equations of our benchmark DSGE model, which is largely taken from Smets and Wouters (2007).[1]

The Representative Household

We consider an economy populated by a representative household with access to financial and physical capital markets so that trading in bonds, investment goods, and state-contingent securities can occur. Household wealth is given by gains from government bonds in nominal per capita terms, B_{t-1}, held at the beginning of period t. Labor income comes from the nominal wage rate, W_t^h, and homogeneous labor, l_t^h, pooled by a set of labor unions, $u \in [0,1]$. Households receive nominal dividends, Π_t^f and Π_t^u, from intermediate producers and labor unions, respectively. Capital

[1] Detailed equations are given in a technical appendix not included here. The appendix and Dynare codes are available on request from the authors.

services incomes are $r_t^K \tilde{K}_t$, where r^K is the real rental price of capital service, $\tilde{K}$.

These revenues are used to pay for consumption, $P_t C_t$, and investment, $P_t I_t$, goods, and for lump-sum taxes expressed in the output price, $P_t T_t$. Moreover, the representative household buys discounted government bonds due at the end of period t, $B_t/(\varepsilon_t^b R_t)$, where ε_t^b is a risk premium shock. Hence, the budget constraint of such a household is given by the following:

$$P_t C_t + P_t I_t + P_t T_t + B_t / \left(\varepsilon_t^b R_t\right)$$
$$\leq W_t^h l_t^h + P_t r_t^K \tilde{K}_t + \Pi_t^f + \Pi_t^u + B_{t-1},$$

which expressed in real terms becomes

$$C_t + I_t + T_t + B_t / \left(\varepsilon_t^b R_t P_t\right) \leq w_t^h l_t^h +$$
$$r_t^K \tilde{K}_t + \left(\Pi_t^f + \Pi_t^u\right) / P_t + B_{t-1} / P_t,$$

where $w_t^h = W_t^h / P_t$ is the real wage received by the household.

Capital services come from the combination of physical capital, K_t, adjusted by capacity utilization, z_t, such that $\tilde{K}_t = z_t K_{t-1}$. Physical capital accumulation implies adjustment cost on investment change, $S(\cdot)$, and the time-varying depreciation process, $\delta(\cdot)$, according to

$$K_t = \left[1 - \delta\left(z_t\right)\right] K_{t-1} + \left(1 - S\left(\varepsilon_t^i \frac{I_t}{I_{t-1}}\right)\right) I_t,$$

where ε_t^i is a shock that deforms the adjustment cost function.[2]

We define the intertemporal utility function as follows:

$$U_t = E_t \left\{ \begin{array}{l} \sum_{j=0}^{+\infty} \beta^j \frac{\left(C_{t+j} - \Theta_{t+j}\right)^{1-\sigma_c}}{1-\sigma_c} \times \\ \exp\left[\eta \varepsilon_t^l \frac{\sigma_c - 1}{1+\sigma_l} \left(l_{t+j}^h\right)^{1+\sigma_l}\right] \end{array} \right\},$$

where σ_c is the intertemporal substitution parameter of consumption, σ_l the intertemporal substitution elasticity of labor, η a scale parameter, and ε_t^l a labor supply shock. External habits are given by $\Theta_t = \theta C_{t-1}$, $0 < \theta < 1$.

The representative household's problem consists of maximizing its intertemporal utility subject to its budget constraint and capital accumulation by choosing the path of C_t, I_t, B_t, z_t, K_t, and l_t^h.

Supply Side

We consider a continuum of intermediate goods producers, $f \in [0,1]$. Each intermediate firm produces a differentiated good used in the production of a final good. Following Kimball (1995), the aggregation function is implicitly given by the following condition:

$$\int_0^1 G^Y\left(\varepsilon_t^y \frac{y_t(f)}{Y_t}\right) df = 1,$$

where $G^Y(\cdot)$ is an increasing, concave function and verifies $G(1) = 1$ and ε_t^y is a shock that distorts the aggregator function. The representative firm in the final good sector maximizes its profit given the prices of intermediate goods, $P_t(f)$, and the price of the final good, P_t.

We assume the following technology in the intermediate producer sector:

$$y_t(f) = \varepsilon_t^a \tilde{K}_t(f)^\alpha \left(\left(1+g\right)^t L_t(f)\right)^{1-\alpha},$$

where ε_t^a is a productivity shock,

$$\ln\left(\varepsilon_t^a\right) = \left(1-\rho_a\right)\ln\left(\bar{\varepsilon}^a\right) + \rho_a \varepsilon_{t-1}^a + v_t^a,$$
$$v_t^a \sim N\left(0, \sigma_a\right),$$

and g is the growth rate of a deterministic, Harrod-neutral technological trend. Assuming that the input markets are perfectly competitive, a firm $f \in [0,1]$ chooses an input mix, $(\tilde{K}_t(f), L_t(f))$, by solving the following program:

$$\min_{\left\{\tilde{K}_t(f), L_t(f)\right\}} w_t L_t(f) + r_t^K \tilde{K}_t(f)$$
$$\text{s.t. } y_t(f) = \varepsilon_t^a \tilde{K}_t(f)^\alpha \left(\left(1+g\right)^t L_t(f)\right)^{1-\alpha},$$

where the real aggregate labor price, w_t, and rental capital rate, r_t^K, are given.

Firms are not allowed to optimally reset their price at each date. With probability $\xi_p > 0$, the

[2] We depart from the Smets and Wouters (2007) model by substituting the initial cost function on change in capital with a time-varying depreciation rate.

firm, f, cannot optimally adjust its price at time t; instead, it follows the following rule:

$$P_t(f) = \bar{\pi}_t^{1-\gamma_p} \pi_{t-1}^{\gamma_p} P_{t-1}(f) = \Gamma_t^p P_{t-1}(f);$$

that is, a nonoptimizing firm sets its price by indexing the current price on a convex combination of past inflation and the inflation target, to be defined subsequently.

The intermediate firm's problem can be written as follows:

$$\max_{\tilde{P}_t(f)} E_t \left\{ \begin{array}{l} \sum_{j=0}^{+\infty} (\beta\xi_p)^j \frac{\lambda_{t+j}}{\lambda_t} \frac{P_t}{P_{t+j}} \\ \left[P_{t+j}(f) - P_{t+j} mc_{t+j} \right] y_{t+j}(f) \end{array} \right\}$$

under conditions

$$P_{t+j}(f) = \Gamma_{t,j}^p \tilde{P}_t(f) \text{ with } \Gamma_{t,j}^p \equiv \begin{cases} \prod_{s=1}^{j} \Gamma_{t+s}^p & \text{if } j > 0 \\ 1 & \text{if } j = 0 \end{cases}$$

and the relative demand function faced by the intermediate firm.

Wage Setting

In this economy, the representative household supplies homogeneous labor, l_t^h, to a unitary continuum of intermediate labor unions indexed by u. Household and labor unions are price takers with regard to the price, W_t^h, of this type of labor, for which the real counterpart corresponds to the marginal rate of substitution of consumption for leisure. The intermediate labor unions aim at differentiating the household's labor and sell this outcome, $l_t(u)$, to a labor agency, setting its price, $W_t(u)$, according to a mechanism à la Calvo (1983). Then the labor agency aggregates these differentiated labor services into a labor package, L_t, and supplies it to productive firms.

Consequently, we assume that the labor agency offers a labor aggregate, L_t, to intermediate firms, derived from differentiated labor unions, $l_t(i)$, according to

$$\int_0^1 G^L\left(\varepsilon_t^w \frac{l_t(i)}{L_t} \right) \mathrm{d}i = 1,$$

where $G^L(\cdot)$ is an increasing, concave function and verifies $G^L(1) = 1$ and ε_t^w a stochastic shock that distorts the aggregator function. Hence, the labor agency maximizes its profit given by

$$\Pi_t = W_t L_t - \int_0^1 W_t(i) l_t(i) \mathrm{d}i.$$

Then, the labor unions set their prices following a Calvo scheme, facing the previous relative demand function and given the wage rate paid to households, W_t^h. More precisely, each labor union seeks to maximize its discounted cash flows by setting the wage rate, $\tilde{W}_t(u)$. With probability ξ_w, the union cannot optimally adjust its wage rate at time t; instead, the union adjusts the wage from consumer price inflation according to the following rule:

$$W_t(u) = \bar{\pi}_t^{1-\gamma_w} \pi_{t-1}^{\gamma_w} (1+g) W_{t-1}(u) = \Gamma_t^w W_{t-1}(u).$$

With probability $1 - \xi_w$, the union is able to choose the optimal wage $\tilde{W}_t(u)$. The labor union's problem can be written as follows:

$$\max_{\tilde{W}(u)} E_t \left\{ \begin{array}{l} \sum_{j=0}^{+\infty} (\beta\xi_w)^j \frac{\lambda_{t+j}}{\lambda_t} \frac{P_t}{P_{t+j}} \\ \left[W_{t+j}(u) - W_{t+j}^h \right] l_{t+j}(u) \end{array} \right\}$$

with the following condition:

$$W_{t+j}(f) = \Gamma_{t,j}^w \tilde{W}_t(f)$$
$$\text{with } \Gamma_{t,j}^w \equiv \begin{cases} \prod_{s=1}^{j} \Gamma_{t+s}^w & \text{if } j > 0 \\ 1 & \text{if } j = 0 \end{cases}$$

and subject to the relative demand function faced by the labor union.

Government, Nominal Distortions, and Aggregation

We assume that government bonds and transfers evolve according to

$$P_t T_t + B_t / \left(\varepsilon_t^b R_t \right) = B_{t-1} + P_t G_t,$$

where G_t is an exogenous process such that the ratio $G/Y = \varepsilon_t^g$ follows an AR(1) process in log.[3] In addition, the central bank sets the current inter-

[3] We use the terms "government shocks" and "external shocks" interchangeably in the following text.

est rate according to the following Taylor rule in its nonlinear form:

$$R_t = \bar{R}^{1-\rho_r} R_{t-1}^{\rho_r} \left[\left(\frac{\pi_t}{\bar{\pi}_t} \right)^{\phi_\pi} \left(\frac{Y_t}{Y_t^{DSGE}} \right)^{\phi_y} \right]^{1-\rho_r} \times$$

$$\left(\frac{Y_t}{Y_{t-1}} \frac{Y_{t-1}^{DSGE}}{Y_t^{DSGE}} \right)^{r_{\Delta y}} \left(\frac{\pi_t}{\pi_{t-1}} \right)^{r_{\Delta \pi}} \varepsilon_t^m,$$

where ρ_r represents the central bank's preference for a smooth interest rate, ε_t^m is a monetary shock, $\bar{\pi}_t$ is a time-varying inflation target, and Y_t^{DSGE} is the output given by a fictional world without nominal rigidities, that is, by setting ξ_p and ξ_w to zero. Hence, this is a measure of the potential output of such a fictional economy.

Despite the heterogeneity of the wages and prices due to the Calvo scheme, we are able to define aggregates for this economy. In fact, total production, that is, the sum of all productions from intermediate firms, y_t, is a priori different from the aggregate final product, Y_t. Consequently, a price distortion, D_t^p, exists such that $y_t = Y_t D_t^p$. The same considerations apply as for the labor market. Total work effort provided by the representative household is l_t^h. Hence, a wage dispersion exists such that $l_t^h = D_t^w L_t$.[4]

We now close the model by deriving the clearing condition on the final product market. First, we need to compute aggregate dividends from intermediate firms:

$$\Pi_t^f = \int_0^1 \left[P_t(f) - P_t mc_t \right] y_t(f) \mathrm{d}f = P_t Y_t - P_t mc_t y_t.$$

Aggregate dividends from labor unions are

$$\Pi_t^u = \int_0^1 \left[W_t(u) - W_t^h \right] l_t(u) \mathrm{d}u = W_t L_t - W_t^h l_t.$$

Combining these two equations with the household's and government nominal budget constraints, and using the competitive market condition for production inputs, leads to

$$C_t + I_t + G_t = Y_t.$$

[4] In fact, these nominal distortions disappear in a linearized model, as in Smets and Wouters (2007). Nevertheless, we need to deal with these distortions as we plan to simulate the model at the second order.

ESTIMATION AND IMPLEMENTATION OF THE PFA METHOD

In this section, we first present the estimation of the two versions of the model on U.S. and euro-area data. Then we describe how we integrate a potential output measure based on the PFA into the model in a fully consistent manner.

Functional Forms and Stochastic Structure

For estimation and simulation, we choose the following functional forms for investment adjustment costs, time-varying depreciation adapted from Greenwood, Hercowitz, and Huffman (1988), and Kimball aggregators that follow the specifications of Dotsey and King (2005):

$$S(x) = \frac{\left(1 - x/(1+g)\right)^2}{2\varphi}, \varphi > 0$$

$$\delta(z) = \psi_1 + \psi_2 \frac{z^d}{d}$$

$$G^i(x) = \frac{1}{(1+\omega_i)\varsigma_i} \left[(1+\omega_i) x - \omega_i \right]^{\varsigma_i} + \left[1 - \frac{1}{(1+\omega_i)\varsigma_i} \right], i \in \{Y, L\}.$$

We choose the following stochastic structure for the exogenous processes in this model:

$$\ln\left(\varepsilon_t^\kappa\right) = \rho_\kappa \ln\left(\varepsilon_{t-1}^\kappa\right) + v_t^\kappa, \; v_t^\kappa \sim N(0, \sigma_\kappa),$$

with $\kappa \in \{i, m, p, b, l, y, w\}$. Finally, we assume that the central bank's target and government expenses on production ratio evolve according to

$$\bar{\pi}_t = \bar{\pi}^{1-\rho_\pi} \bar{\pi}_{t-1}^{\rho_\pi} \varepsilon_t^p$$

$$\ln\left(\varepsilon_t^g\right) = \left(1 - \rho_g\right) \ln\left(\bar{g}_y\right) + \rho_g \ln\left(\varepsilon_{t-1}^g\right) + v_t^g,$$

$$v_t^g \sim N\left(0, \sigma_g\right).$$

Then, before starting estimation procedures, we need to make the model stationary. Indeed, as the model features a balanced growth trend, it is necessary to turn it into its intensive form for simulations. All real variables of interest are deflated

by the deterministic trend $(1 + g)^t$. We then rewrite the model's equations with intensive variables.[5]

Priors Distributions, Calibration, and Data

We use Bayesian techniques to estimate the main free parameters of the model. Broadly speaking, we compute by numerical simulation the maximum of the posterior density of the parameters by confronting a priori knowledge about them, through the likelihood function, against data.[6] The first column of Table 1 shows the different priors set to estimate both the U.S. and euro-area models.

Almost all of the model's parameters are estimated, with the following exceptions: The time preference parameter β is set at 0.998; the Kimball function's parameter ζ_Y and ζ_L is calibrated at 1.02 as in Dotsey and King (2005); and the average quarterly growth rate of gross domestic product (GDP), g, is set at 0.66 percent for the euro area and 0.37 percent for the U.S. economy, based on our database. Note that the prior density functions are quite noninformative for most of the estimated parameters except for inertia coefficients of productivity shocks and what we call "government shocks." We used the previous result of highly persistent shocks in previous works as a prior belief (e.g., Smets and Wouters, 2007).

The data sources are as follows. We use time series from 1970:Q1 to 2007:Q4 (United States) and 2006:Q4 (euro area). For U.S. data, the GDP, consumption, investment, and GDP deflator are from the Bureau of Economic Analysis national accounts. The capacity utilization rate and nominal interest rate—the federal funds effective rate—are from the Federal Reserve Board database. For the euro-area data, GDP, consumption, investment, short-term interest rate, and GDP deflator are from the Area-wide Model (AWM) database (Fagan, Henry, and Mestre, 2001). Capacity utilization rate data are from the Eurostat database. Finally, data on labor markets have been used to detrend extensive variables. Total U.S. employment and hours worked for the U.S. and euro-area economies are from the OECD's Economic Outlook database (OECD, 2005). European total employment data are from the AWM database. All extensive variables, namely GDP, consumption, and investment, are first detrended through a Hodrick-Prescott (HP) filter with parameter 1600 using a trend in labor that consists of total hours worked. Then, these variables are deflated by the GDP deflator. We therefore compute the average quarterly growth rate of real gross productivity and detrend again all extensive variables by the corresponding deterministic time trend. Finally, these variables are divided by the mean of GDP over the period.[7]

Implementing the Production Function Method

The first step consists of estimating the benchmark DSGE model for the two economies and checking the consistency of estimates given by the two last columns of Table 1.[8] We then simulate the model to obtain consistent time series for production, investment, labor, and capacity utilization.

We now are able to (i) compute the physical capital stock series according to the permanent inventory method (PIM) and, taking into account the deterministic trend,

$$k_t^{PIM} = \frac{1-\bar{\delta}}{1+g} k_{t-1}^{PIM} + i_t,$$

as well as the age of capital,

$$age_t = \frac{1-\bar{\delta}}{1+g} \frac{k_{t-1}^{PIM}}{k_t^{PIM}} \left(age_{t-1} + 1\right), \text{ and}$$

(ii) extract the Solow residual, s_t, as

$$s_t = \ln\left(y_t\right) - \alpha \ln\left(k_t^{PIM}\right) - (1-\alpha)\ln(L).$$

[5] As written in the technical appendix.

[6] See Schorfheide (2000) and Smets and Wouters (2003).

[7] We deliberately exclude data on wages and labor in the estimation process primarily because of the lack of labor market sophistication in the model.

[8] As a consistency check, one can verify that the posterior modes obtained by the estimation process correspond to the maximum of the likelihood function in the parameter direction. Such representations are given in the technical appendix.

Table 1

Priors Distributions and Posterior Modes

	Prior distribution			Posterior modes	
Parameter	Type	Mean	SD	Euro area	United States
Preferences					
θ	beta	0.500	0.2000	0.3210	0.2248
σ_c	norm	1.500	0.5000	1.1592	1.4674
σ_l	norm	2.000	0.5000	1.9061	0.6218
Production and technology					
d	gamma	1.500	0.2000	1.6581	1.8098
$\bar{\delta}$	beta	0.500	0.2000	0.0820	0.0569
ϕ	norm	5.500	5.0000	0.1824	0.0890
α	beta	0.500	0.2000	0.2409	0.1907
Kimball aggregators					
ω_Y	norm	–6.000	5.0000	–5.1063	–4.4232
ω_L	norm	–18.000	5.0000	–16.0926	–16.1249
Calvo settings					
ξ_p	beta	0.500	0.2000	0.5411	0.4252
γ_p	beta	0.500	0.2000	0.0432	0.1295
ξ_w	beta	0.500	0.2000	0.4431	0.5100
γ_w	beta	0.500	0.2000	0.7980	0.7781
Steady-state values					
$\bar{z}$	norm	0.814	0.1000	0.8255	0.8040
$\bar{g}_y$	beta	0.200	0.1000	0.1935	0.0461
$\bar{\pi}$	norm	1.014	0.1000	1.0077	1.0125
$\bar{L}$	norm	1.000	0.1000	1.0002	1.0001
$\bar{y}$	norm	1.000	0.1000	0.8602	0.9178
Autoregressive parameter of shocks					
ρ_a	beta	0.990	0.0010	0.9908	0.9904
ρ_i	beta	0.500	0.2000	0.1739	0.1655
ρ_π	beta	0.500	0.2000	0.0510	0.0961
ρ_m	beta	0.500	0.2000	0.9686	0.9391
ρ_p	beta	0.500	0.2000	0.0510	0.0960
ρ_l	beta	0.500	0.2000	0.4999	0.4989
ρ_g	beta	0.970	0.0100	0.9750	0.9980
ρ_b	beta	0.500	0.2000	0.9662	0.9571
ρ_w	beta	0.500	0.2000	0.5007	0.4998
ρ_y	beta	0.900	0.0500	0.8473	0.9180

Table 1, cont'd

Priors Distributions and Posterior Modes

Parameter	Prior distribution			Posterior modes	
	Type	Mean	SD	Euro area	United States
Taylor rule					
ϕ_π	norm	2.000	0.5000	2.5860	2.6345
ϕ_y	norm	0.100	0.0500	0.0905	0.1394
$r_{\Delta y}$	norm	0.000	0.5000	0.3419	0.6927
$r_{\Delta\pi}$	norm	0.300	0.1000	0.1384	0.1252
ρ_r	beta	0.500	0.2000	0.9268	0.8505
Standard deviation of shocks					
v_i	invg	0.100	2.0000	0.0208	0.0512
v_a	invg	0.010	2.0000	0.0072	0.0061
v_p	invg	0.001	2.0000	0.0033	0.0026
v_m	invg	0.001	2.0000	0.0004	0.0006
v_b	invg	0.100	2.0000	0.0018	0.0020
v_l	invg	0.001	2.0000	0.0005	0.0005
v_g	invg	0.001	2.0000	0.0197	0.0786
v_w	invg	0.001	2.0000	0.0005	0.0005
v_y	invg	0.100	2.0000	0.0184	0.0209
Data density				3,634	3,598

NOTE: This table shows prior distribution of the benchmark model parameters and estimation results at the mode of the marginal density posteriors. Prior probability density functions are normal (norm), beta (beta), or inverse gamma (invg). SD, standard deviation.

Table 2

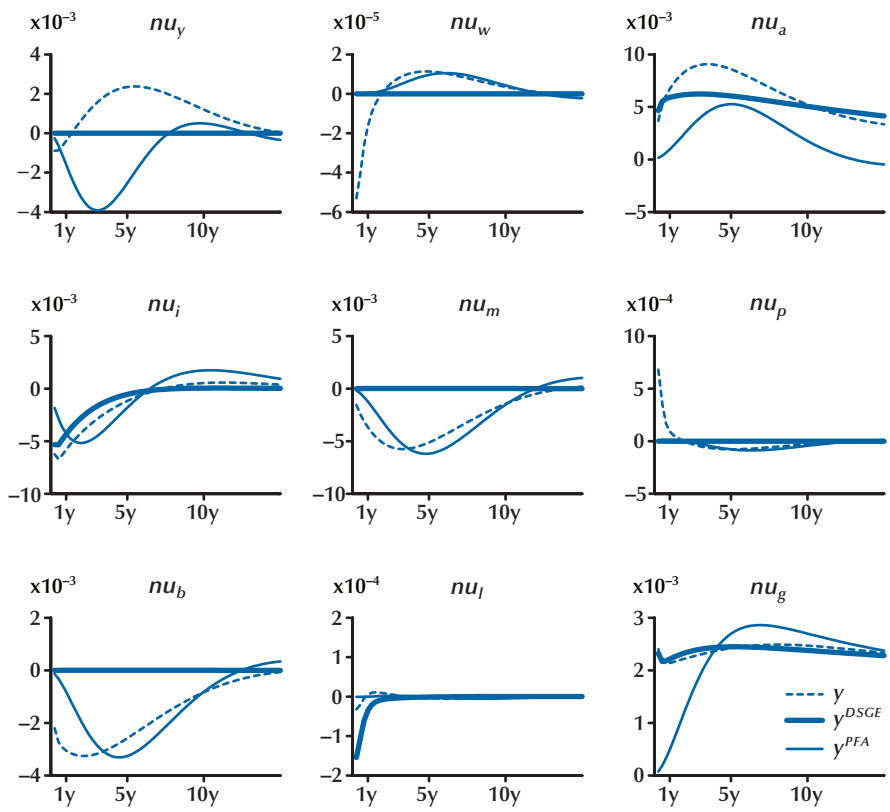
Results Estimates of TFP Equation

Study area	γ_0 intercept	γ_1 s_{t-1}	γ_2 $\ln(z_t)$	γ_3 age_t	R^2
Euro area	–0.0100	0.9059	–0.1226	–4.3e-03	0.9974
	(0.0061)	(0.0081)	(0.0123)	(7.0e-04)	—
United States	–0.0300	0.8784	–0.1477	–2.2e-03	0.9946
	(0.0082)	(0.0090)	(0.0123)	(5.8e-04)	—

NOTE: This table shows results estimates of the TFP equation based on simulated series of 3,000 occurrences, where the first 1,000 have been dropped. We made 1,000 regressions. The figures in the table correspond to the average parameters over these regressions. Average standard deviations are listed in parentheses.

Figure 1

Impulse Response Function for Production and DSGE/PFA Measures of Potential Output (United States)



Finally, we estimate the following TFP equation:

$$s_t = \gamma_0 + \gamma_1 s_{t-1} + \gamma_2 \ln(z_t) + \gamma_3 age_t + \varepsilon_t,$$

where ε_t is an i.i.d. process.[9]

Table 2 gives the estimates of the TFP equation for both the U.S. and euro-area model economies. It is worth noting that results show a negative coefficient on the capacity utilization, contrary to what we assumed as an economic intuition in the section on benchmarking the DSGE model.

We then compute the potential production based on the PFA. First, we assume that potential capital is taken as k_t^{PIM}, as computed from the PIM. Then, we use filtered data to assess potential employment, L_t^{Filt}.[10] Finally, we define the medium-term potential TFP, $\hat{s}_t^{MT}$, from our previous estimates by setting $z_t \equiv \overline{z}$ and eliminating the lagged term[11]:

[9] See Cahn and Saint-Guilhem (2009).

[10] More specifically, we use a moving average version of the HP filter—formally, if a process, x_t, can be split between a cyclical part, c_t, and a smooth trend, m_t. The HP filter defines the cyclical part as

$$c_t = \frac{\lambda(1-\mathfrak{L})^2\left(1-\mathfrak{L}^{-1}\right)^2}{1+\lambda(1-\mathfrak{L})^2\left(1-\mathfrak{L}^{-1}\right)^2} x_t,$$

where $\mathfrak{L}$ is the lag operator. Expanding this expression and considering that $c_t = x_t - m_t$, we use the following relation to define potential labor:

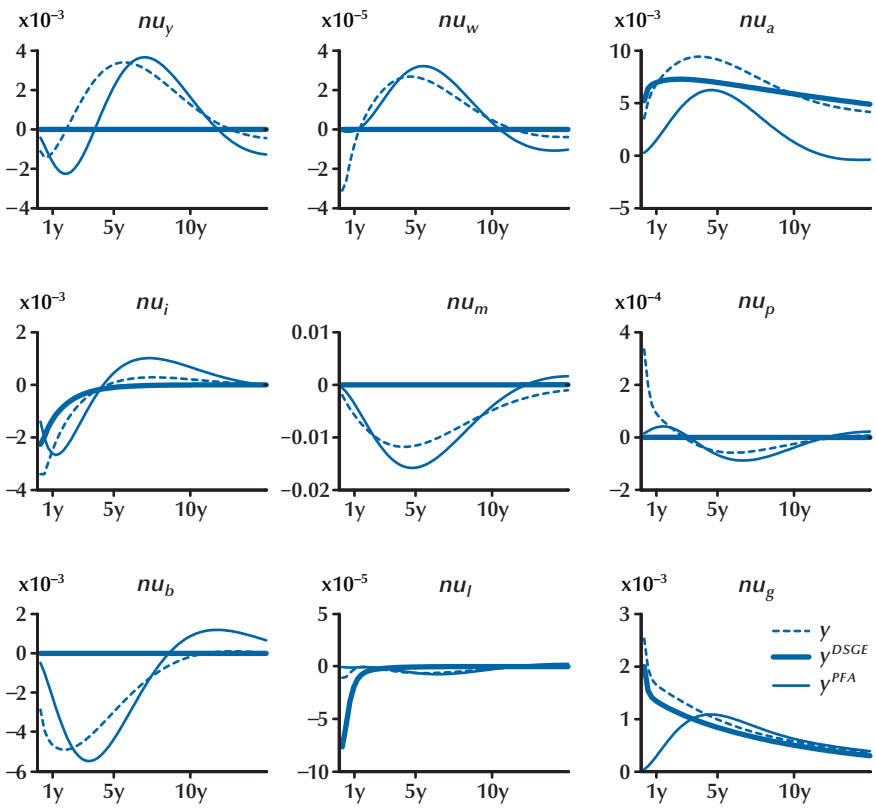
$$L_t = L_t^{Filt} + \lambda\left(\mathfrak{L}^{-2} - 4\mathfrak{L}^{-1} + 6I - 4\mathfrak{L} + \mathfrak{L}^2\right)L_t^{Filt}.$$

Finally, we set $\lambda = 1600$ as is standard for quarterly economic time series.

[11] See Cahn and Saint-Guilhem (2009).

Figure 2

Impulse Response Function for Production and DSGE/PFA Measures of Potential Output (Euro Area)



$$\hat{s}_t^{MT} = \frac{\gamma_0}{1-\gamma_1} + \frac{\gamma_2}{1-\gamma_1}\ln(\bar{z}) + \frac{\gamma_0}{1-\gamma_1} age_{t+1}.$$

Consequently, the potential output based on our production function method is given by

$$y_t^{PFA} = e^{\hat{s}_t^{MT}} \left(k_t^{PIM}\right)^{\alpha} \left(L_t^{Filt}\right)^{1-\alpha}.$$

DISCUSSION

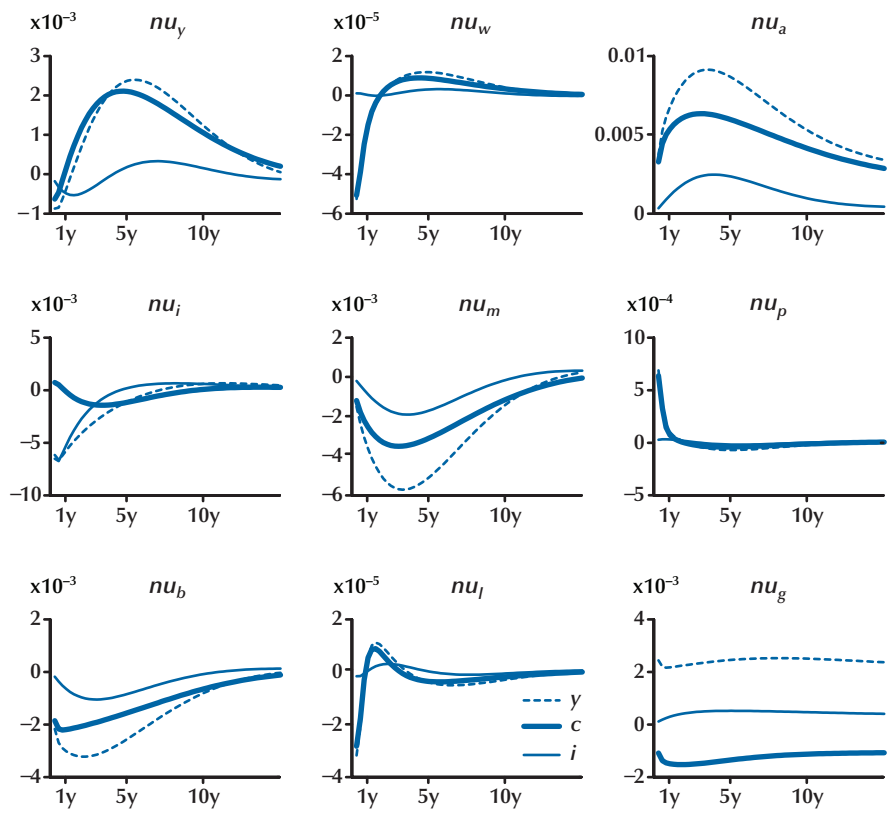
In this section, we analyze and compare the dynamic behavior of the DSGE and PFA estimates of potential output through impulse response functions (IRFs) and variance decomposition. In the following, the terms "U.S. economy/model" and "euro-area economy/model" refer to the models estimated on U.S. data or euro-area data, respectively.

IRF Analysis

Figures 1 and 2 plot the IRFs of the stochastic shocks for actual output and both DGSE- and PFA-based measures of potential output, calculated with the estimated parameters given in Table 1. Figures 3 through 8 show the IRFs for various factors (output, consumption, investment, nominal interest rate, inflation, and DSGE- and PFA-based output gaps). Results for U.S. and euro-area models are broadly similar, except for the response to an external expenses shock, v^g, for which the U.S. response appears to be much more inert than the euro-area responses. The fol-

Figure 3

Impulse Response Function for Output, Consumption, and Investment (United States)



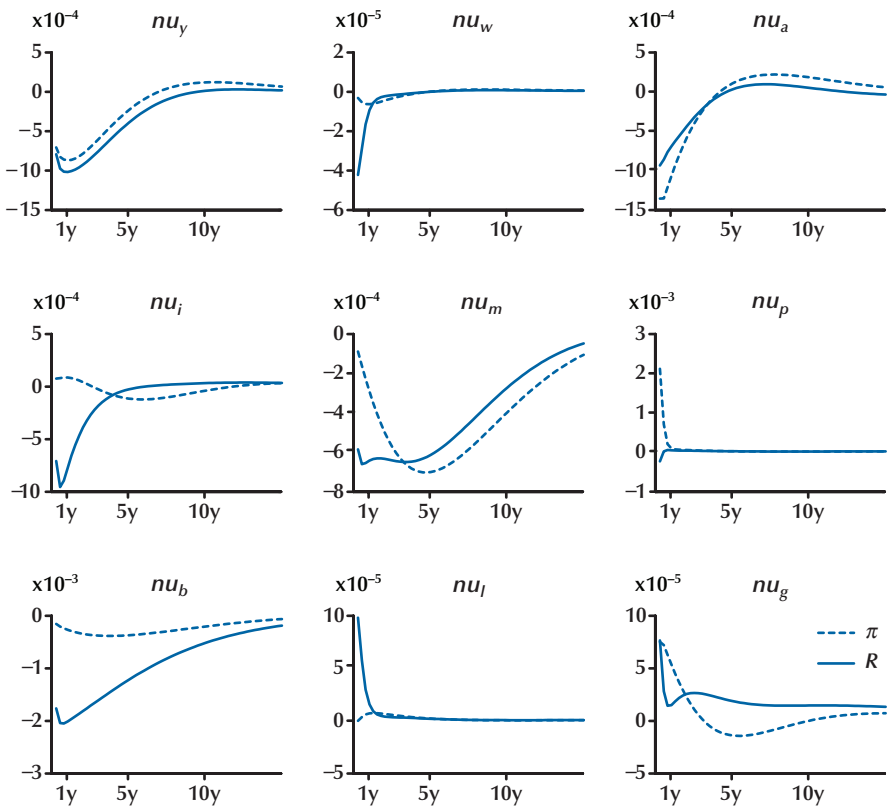
lowing analysis applies for both economies, apart from this shock.

The figures show that after a positive productivity shock, v^a, actual, DSGE, and PFA potential outputs rise together, but the PFA measure rises more gradually. Moreover, the response of the model-based potential output seems to be more persistent for the DSGE than for the PFA. Indeed, a positive productivity shock results in an increase in investment, and therefore the age of capital stock grows gradually, as do the medium-term TFP and PFA potential outputs. On the other hand, the productivity shock instantly affects both the productivity term and the Solow residual. Consequently, after such a shock, both actual and PFA potential outputs evolve similarly, but the gap between them remains constant for a longer time than with the DSGE potential output (see Figures 5 and 8).

The effect of a positive—quantitatively negative in its effect—investment shock, v^i, leads to similar dynamics for the three output measures. The shock deforms the adjustment cost function, leading to an increase in the cost of new capital. Hence, investment falls and capital stock shows a hump-shaped decrease, reflected in its age and then in potential TFP. Interestingly, all these variables cross their steady-state path simultaneously after about 6 years. Before, the PFA potential output is below the DSGE potential output, and this order changes after the date; the actual output lies between the two measures. This implies that the two related gap measures evolve in opposite directions after an investment shock.

Figure 4

Impulse Response Function for Nominal Interest Rate and Inflation (United States)



With respect to a positive labor supply shock, v^l, PFA potential output does not react, whereas DSGE potential output decreases instantaneously, as does actual output but to a lesser extent. In fact, labor in the world without nominal rigidities can adjust more rapidly, and the reaction of DSGE potential output is one order of magnitude higher than for actual and PFA potential outputs.

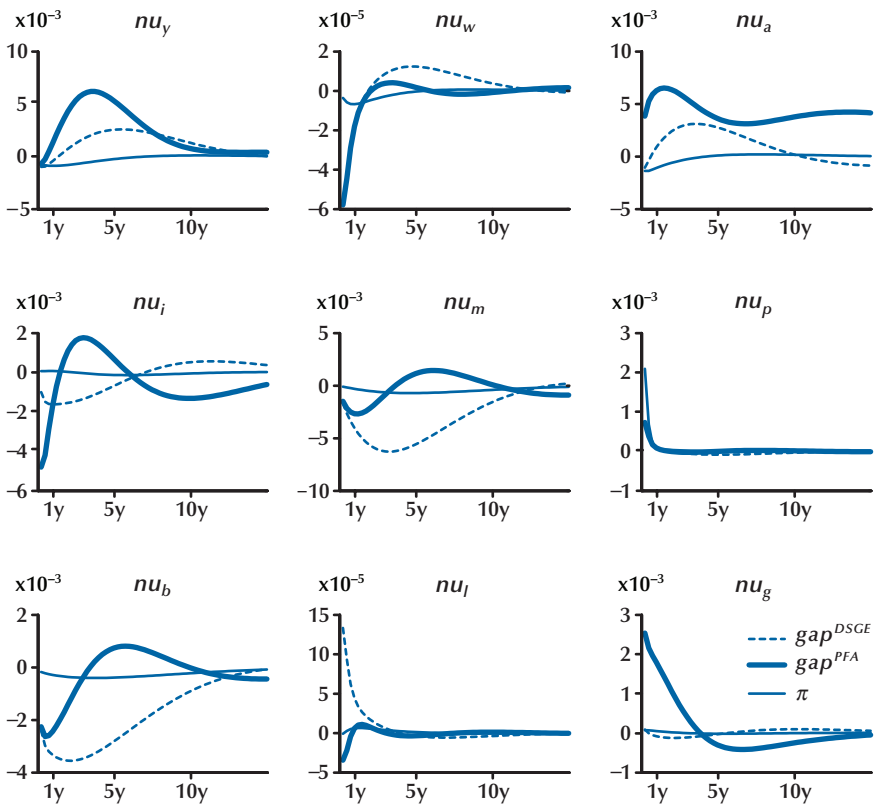
Conversely, after a positive government shock, v^g, actual and DSGE potential outputs shift upward suddenly, whereas PFA potential output gradually reaches their level. After the shock, demand for output shifts upward instantly, coinciding with a higher level of employment. Hence, potential employment grows gradually and then results in the slower increase in PFA-based potential output. Note that the response to the government shock of the U.S. model is more persistent than for the euro-area ones. This is mainly due to a more persistent stochastic structure of the shock estimated for the United States.

Not surprisingly, DSGE potential output does not respond to any nominal shocks—namely, markup, monetary, or equity premium shocks—as these shocks do not enter into the real side model. The most remarkable fact is that PFA potential output reacts significantly to such shocks. After a positive monetary shock to the interest rate, v^m, both actual and PFA output show a hump-shaped decrease. The qualitative effects of the equity premium shock, v^b, are quite similar.

The model economies show similar responses to the price and wage markup shocks with first an instantaneous fall in actual output and then a

Figure 5

Impulse Response Function for Inflation-, DSGE-, and PFA-Based Output Gaps (United States)



hump-shaped increase. Nevertheless, the actual output reaction to a wage markup shock is about two orders of magnitude less than the response to a price markup shock. PFA potential output responds to these shocks in a similar manner but more gradually, generating a persistent drift.

Variance Decomposition

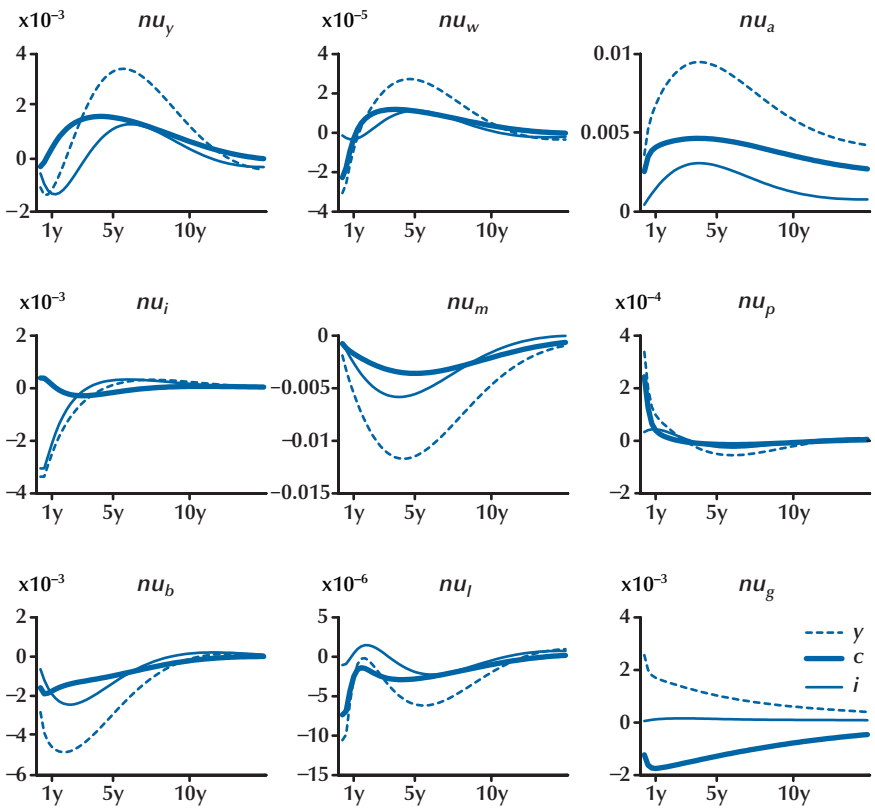
Table 3 shows the contribution of each structural shock to the asymptotic forecast error variance of the endogenous variables shown in Table 4. For the U.S. economy, the productivity shock seems to dominate asymptotically the sources of actual and DSGE-based potential outputs by about 50 percent and 60 percent, respectively. A government spending shock is the other main source of fluctuations, accounting for 27 percent of actual production and 37 percent of DSGE potential.

The interest rate shock appears to create the most striking difference between actual and DSGE potential output variance—it amounts to about 55 percent of the related output gap measure. For the PFA potential measure, the external spending shock accounts for 43 percent of the variance as the main contributor. The productivity shock contribution reaches only 15 percent, less than the interest shock (21 percent). All in all, contrary to the DSGE-based measure, the productivity shock accounts for 68 percent of the PFA output gap variance.

For the euro-area model economy, the variance decomposition of actual production is quite

Figure 6

Impulse Response Function for Output, Consumption, and Investment (Euro Area)



similar to that of the United States. Nevertheless, almost all the variance of DSGE-based potential output seems to be derived from the productivity shock, whereas the largest part of the PFA potential output variance comes from the interest shock (76 percent) and productivity shock to a lesser extent (12 percent). As a result, DSGE output gap variations come primarily from the interest shock (82 percent), and PFA gap variance is derived from the productivity shock (74 percent).

Finally, Table 4 shows that both DSGE and PFA potential growth are less volatile than actual output. Nevertheless, one could not conclude that the PFA-based measure is smoother than the DSGE one.

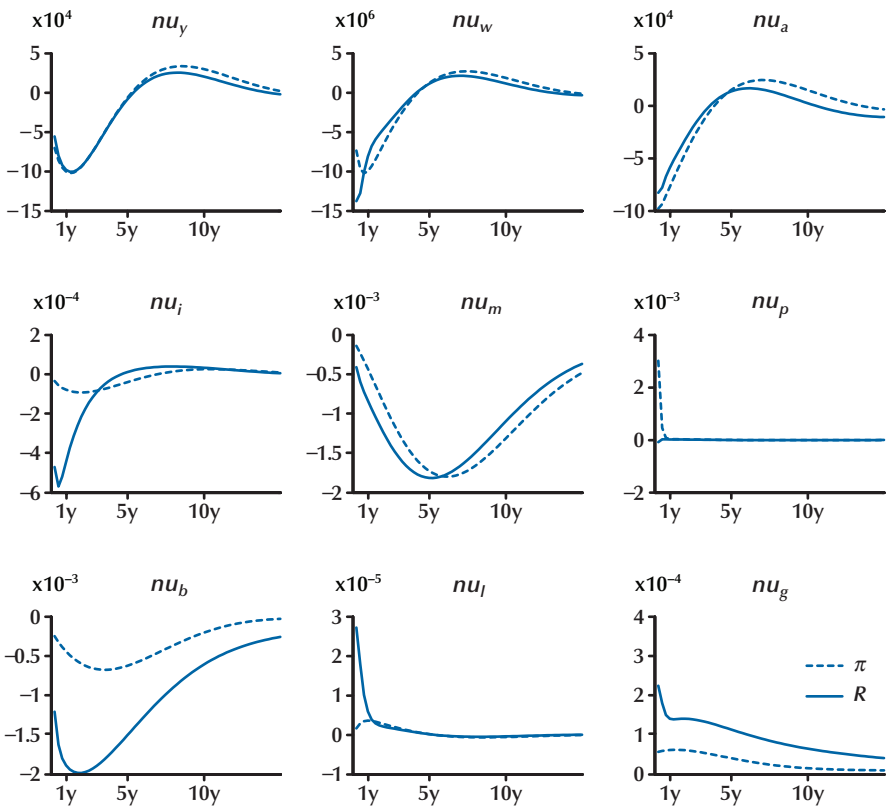
Implications

Our analysis suggests that the PFA and the DSGE approaches to potential output measurement differ significantly, at least from a business cycle perspective. For two different models—one close to the U.S. data, the other to the euro-area data—the output gap related to the DSGE measure captures mainly nominal shocks, which in summation amounts to more than 80 percent (about 97 percent for the U.S. model) of the gap variance. Alternatively, the PFA gap reacts mainly to productivity shock (about 70 percent of the variance.)

As a consequence, it seems to us that using the PFA to compute potential output and the related output gap presents some drawbacks related to

Figure 7

Impulse Response Function for Nominal Interest Rate and Inflation (Euro Area)

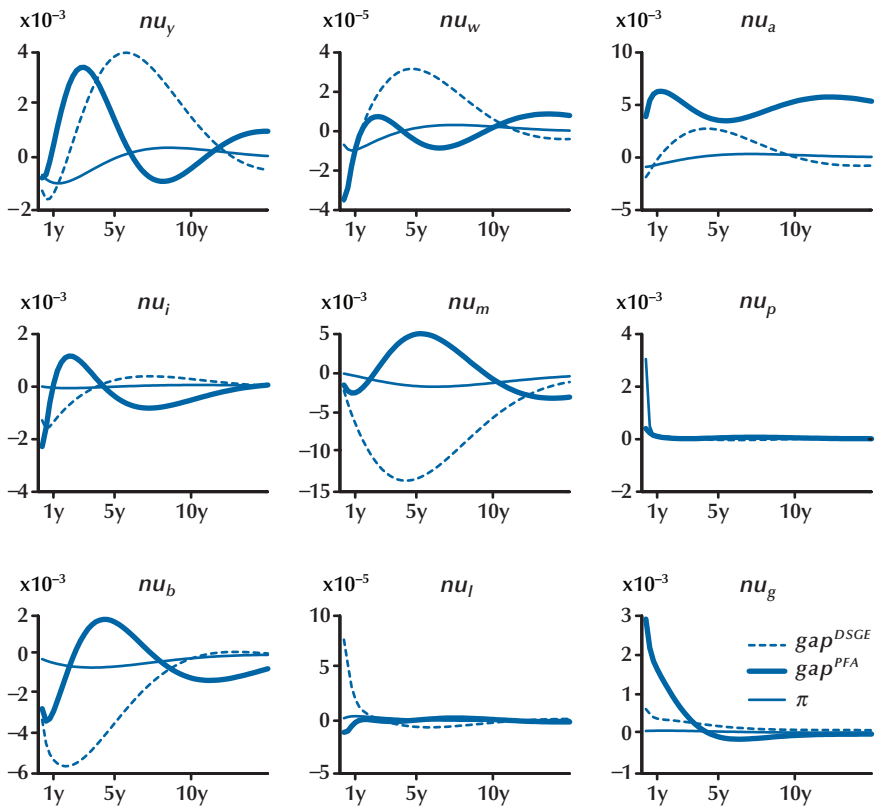


its ability to properly reflect inflationary pressures related to nominal shocks. In contrast, the DSGE-based potential output measure could lead to misstatements about potential growth as this measure reacts to temporary but persistent shocks such as productivity shocks. These two assessments can lead to contradictions in terms of economic diagnostics. For instance, assuming that the model is the one that generates the actual data, one could think that during the 1990-95 period, GDP growth in the United States (2.4 percent) was below its potential based on the DSGE measure (2.7 percent), as stated in Table 5. One would reach an opposing conclusion using the PFA-based measure (1.7 percent). The same contradiction holds for the euro-area economy during the 2000-05 period.

From an empirical point of view, these results tend to moderate the possible structural interpretations of the international comparison based on the PFA. Indeed, if one believes that some structural shocks drive the dynamics of economic variables and wants to compare potential growth of several economies using the PFA, the fact that the results depend on the idiosyncratic shocks faced by each economy must be considered. Consequently, this argues for a normalization of such structural shocks before applying the PFA. For instance, based on the PFA (left side of Table 5), it appears that actual growth in the euro-area economy stood below its PFA potential in the past 15 years. Conversely, the U.S. economy's actual growth was above its PFA potential. Moreover, the U.S. PFA potential was higher than the

Figure 8

Impulse Response Function for Inflation-, DSGE-, and PFA-Based Output Gaps (Euro Area)



euro area's. Does it clarify the need for structural reforms in the European economy to keep pace with the U.S. economy? Imagine that both economies interchanged the structural shocks they faced. Would we observe identical behavior? The three right columns of Table 5 present the results of such an experiment; they lead to the exact opposite conclusion regarding the comparison between the United States and the euro area.

Alternatively, a monetary authority that must conduct interest rate policy based on a Taylor rule that includes an output gap measure could make the opposite decision depending on the method used to measure drift between actual and potential output. For instance, after a positive productivity shock, the central bank could decide to instantaneously increase the nominal interest rate if based on the PFA gap estimates, whereas the decision would be to decrease the interest rate (as shown in Figure 4), at least in a DSGE framework.

CONCLUSION

In this article, we compared the PFA measure of potential output with the DSGE definition of potential output in a fully integrated framework. We estimated a DSGE model for U.S. and euro-area data and integrated into the two versions of the model a PFA measure of potential output fully consistent with the model. Results have shown that, in a DSGE framework, the PFA leads to potential output measures that are not exempt from

Table 3

Variance Decomposition

Shocks	y_t	y_t^{DSGE}	y_t^{PFA}	π_t	c_t	i_t	R_t	gap_t^{DSGE}	gap_t^{PFA}
United States									
Productivity shock	51	59	15	24	65	28	4	14	67
Inflation target shock	—	—	—	11	11	—	—	—	—
Labor supply shock	—	—	—	—	—	—	—	—	—
External spending shock	27	37	43	—	15	8	—	—	1
Investment shock	5	4	10	1	1	45	4	3	4
Equity premium shock	4	—	6	11	4	4	68	18	2
Interest rate shock	11	—	21	32	11	14	12	55	4
Price distortion shock	2	—	5	21	4	1	12	10	22
Wage distortion shock	—	—	—	—	—	—	—	—	—
Euro area									
Productivity shock	48	99	12	4	67	22	2	3	73
Inflation target shock	—	—	—	7	—	—	—	—	—
Labor supply shock	—	—		—	—	—	—	—	—
External spending shock	1	1	1	—	5	—	—	—	1
Investment shock	1	—	1	—	—	3	1	—	1
Equity premium shock	5	—	7	8	3	7	44	10	3
Interest rate shock	42	—	75	72	22	65	48	82	18
Price distortion shock	3	—	4	9	3	3	5	5	4
Wage distortion shock	—	—	—	—	—	—	—	—	—

NOTE: This table presents the theoretical variance decomposition among the model's shocks (expressed in percent).

Table 4

Theoretical Moments

	United States		Euro area	
Variable	Mean	SD	Mean	SD
y	0.9178	0.0794	0.8602	0.0909
y^{DSGE}	0.9178	0.0667	0.8602	0.0611
Y^{PFA}	0.9178	0.0620	0.8602	0.0819
C	0.7237	0.0525	0.5083	0.0433
π	1.0125	0.0066	1.0077	0.0117
R	1.0239	0.0105	1.0136	0.0143
i	0.1517	0.0229	0.1855	0.0347
gap^{DSGE}	0.0000	0.0394	0.0000	0.0755
gap^{PFA}	0.0000	0.0534	0.0000	0.0598

Table 5

Annual Potential Growth Comparison

	United States			United States*		
Period	y	y^{DSGE}	y^{PFA}	y	y^{DSGE}	y^{PFA}
1990-1995	2.4	2.7	1.7	2.0	1.6	2.6
1995-2000	3.9	3.4	3.8	2.4	2.5	3.0
2000-2005	2.4	1.9	1.7	0.0	1.2	0.6
1990-2005	2.9	2.6	2.4	1.5	1.8	2.1
	Euro area			**Euro area†**		
	y	y^{DSGE}	y^{PFA}	y	y^{DSGE}	y^{PFA}
1990-1995	1.8	2.5	1.5	2.4	3.6	2.2
1995-2000	2.4	2.8	2.7	4.2	4.1	2.3
2000-2005	1.9	1.9	2.9	4.8	3.9	3.1
1990-2005	2.0	2.4	2.4	3.8	3.8	2.5

NOTE: This table shows actual and potential growth on average over different subperiods. Figures are given in percent. They also include both the deterministic and labor trends. *U.S. model simulated with euro-area smoothed shocks. †Euro-area model simulated with U.S. smoothed shocks.

the effects of nominal or temporary shocks. The empirical implication of these results is that estimates of potential output based on an ad hoc PFA could be highly dependent on transitory phenomena. Moreover, cross-country differences in potential output based on the PFA are likely to reflect not only structural differences, but also different patterns of shocks across time. This leads to the assessment of the quantitative role of shocks in cross-country differences in potential output. One way to address this issue is to implement in a DSGE model a scenario comparing potential output across economies confronted by the same shocks across time, while exhibiting differences in structural parameters.

However, to answer this question in a more satisfactory manner, we need to improve the present study in several directions. First, it would be of particular interest to identify the causes of divergences between PFA and DSGE potential output measures. Such an analysis could be conducted parameter by parameter to assess their weight on the discrepancy between the two assessments. Second, one would need to improve the estimation procedure by identifying the marginal posterior density of the model through Markov-chain Monte Carlo simulations on the one hand, and by allowing structural breaks in the TFP regression equation on the other hand. Finally, one could study the implications for monetary policy of the use of PFA rather than DSGE measures of output gap in a class of central bank decision rules. Obviously, these studies should be performed with an enhanced model, especially with regard to the modeling of the labor market, with an extension of the model introducing unemployment and participation considerations to account for additional sources of fluctuations in potential output and the output gap.

REFERENCES

Cahn, Christophe and Saint-Guilhem, Arthur. "Potential Output Growth in Several Industrialised Countries: A Comparison." *Empirical Economics*, 2009 (forthcoming).

Calvo, Guillermo A. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics*, September 1983, *12*(3), pp. 383-98.

Dotsey, Michael and King, Robert G. "Implications of State-Dependent Pricing for Dynamic Macroeconomic Models." *Journal of Monetary Economics*, January 2005, *52*(1), pp. 213-42.

Edge, Rochelle M.; Kiley, Michael T. and Laforte, Jean-Philippe. "Natural Rate Measures in an Estimated DSGE Model of the U.S. Economy." Finance and Economics Discussion Series No. 2007-08, Federal Reserve Board, Washington, DC; www.federalreserve.gov/pubs/feds/2007/200708/200708pap.pdf.

Fagan, Gabriel; Henry, Jerome and Mestre, Ricardo. "An Area Wide Model (AWM) for the Euro Area." ECB Working Paper No. 42, European Central Bank, January 2001; www.ecb.int/pub/pdf/scpwps/ecbwp042.pdf.

Greenwood, Jeremy; Hercowitz, Zvi and Huffman, Gregory W. "Investment, Capacity Utilization, and the Real Business Cycle." *American Economic Review*, June 1988, *78*(3), pp. 402-17.

International Monetary Fund. "France—2007 Article IV Consultation Concluding Statement." International Monetary Fund, November 19, 2007; www.imf.org/external/np/ms/2007/111907.htm.

Kimball, Miles. "The Quantitative Analytics of the Basic Neomonetarist Model." *Journal of Money, Credit, and Banking*, 1995, *27*(4 Part 2), pp. 1241-77.

Neiss, Katherine and Nelson, Edward. "Inflation Dynamics, Marginal Costs and the Output Gap: Evidence from Three Countries." *Journal of Money, Credit, and Banking*, December 2005, *37*(6), pp. 1019-45.

Organisation for Economic Co-operation and Development. *OECD Economic Outlook No. 78*. December 2005.

Schorfheide, Frank. "Loss Function-Based Evaluation of DSGE Models." *Journal of Applied Econometrics*, *15*(6), pp. 645-70.

Smets, Frank and Wouters, Rafael. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association*, 2003, *1*(5), pp. 1123-75.

Smets, Frank and Wouters, Rafael. "Shocks and Frictions in U.S. Business Cycles: A Bayesian DSGE Approach." *American Economic Review*, June 2007, *97*(3), pp. 586-606.

Commentary

Jon Faust

The Economic Policy conference at the Federal Reserve Bank of St. Louis has for several decades been one of the premier monetary policy conferences worldwide, and it is a great privilege to participate in this conference focusing on the measurement and forecasting of potential growth. I am particularly pleased to be discussing the paper by Christophe Cahn and Arthur Saint-Guilhem (2009), which is a beautiful example of a broad class of work that explores how traditional economic concepts and measures relate to similar concepts in the context of dynamic stochastic general equilibrium (DSGE) models that are rapidly coming into the policy process. This class of work is vitally important if policymakers are to meld successfully traditional methods and wisdom with the new models to improve the policy process. I will mainly attempt to explain this class of work, why it is important, and some techniques for improving it. While my points are fairly generic, the Cahn–Saint-Guilhem paper provides an excellent case study for illustrating the key issues.

DSGE MODELS AND A NEW CLASS OF RESEARCH

Around 1980, Lucas, Sims, and others issued devastating critiques of existing monetary policy models. One basis for these critiques was the claim that the existing methods were substantially ad hoc relative to the ideal to which the profession should aspire. While this critique was undeniably valid, the absence of better-founded alternatives meant that more-or-less traditional ad hoc approaches continued to be used and refined at central banks for the next 25 years or so. Meanwhile, the profession did the basic research required to create models with sounder foundations.

In the past few years, DSGE models have advanced to the point that they are coming into widespread use at central banks around the world. These models are still rife with ad hoc elements, but there is no doubt that there has been an order of magnitude advance in the interpretability of the predictions of the model in terms of well-articulated economic theory.

There is still considerable disagreement, however, over the degree to which the new models should supplant the traditional methods. I do not want to argue this point. Rather, I want to assert that these models have at least advanced to the point that they constitute interesting laboratories in which to explore various claims and principles that are important in the policy process. My focus is on how the models can best play this role.

Consider an analogy to medical research. In attempting to understand the toxicology of drugs in humans, we often use animal models. That is, we check if the drug kills the rat before we give it to humans. In any given pharmacological context, there is generally substantial disagreement on how literally we should take the model when extrapolating the results to humans. Despite this

Jon Faust is the Louis J. Maccini Professor of Economics at Johns Hopkins University.

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 241-46.
© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

disagreement, there is a broad consensus that the rat model is extremely valuable in formulating policy.

Similarly, I think we should all agree that DSGE models have at least attained something akin to rat, or at least fruit fly, status. Under this agreement, a wide range of work becomes valuable and important. In particular, I think we should aggressively explore basic macroeconomic propositions treating these model economies as interesting economic organisms.

Although I am not sure the authors view it this way, the Cahn–Saint-Guilhem paper can be viewed in this perspective. Some notion of potential output is often at the center of policy discussions. One traditional measure of potential is based on the production function approach (PFA) as clearly described in the paper. Analysis of optimal policy in DSGE models suggests that for some purposes we should focus on a concept of potential as measured by the *efficient* level of output, known as flexible price output (FPO). FPO potential measures what output would be if certain distortions were not present.[1]

If we are to smoothly and coherently bring DSGE models and the associated measures into the policy process, it is important to know how PFA and FPO potential relate in the real world. One very useful step in this process, I argue, is exploring how both concepts operate in the simpler context of the DSGE model. That is, first understand the concepts as fully as possible in the rat before moving to the human case.

This type of work is relatively straightforward conceptually. Broadly, we must specify how to compute a model-based analog of both PFA and FPO potential. Then we simulate a zillion samples from the model, calculate both measures on each, and then summarize apparent similarities and dissimilarities.[2] For example, we might ask whether our traditional interpretation of PFA potential is correct in the context of the DSGE model.

The paper focuses on a particular question of this type. Movements in PFA potential are, in practice, often attributed to medium-term "structural features" of the economy as opposed to transitory demand or supply features. Is the interpretation warranted in the DSGE model? The paper finds (see their Table 3) that it is not. That is, a large portion of the variance of PFA potential is attributable to factors we would not usually consider "structural" in the sense this term is used in these discussions. FPO potential looks more structural in this regard. The paper elaborates this key result in a number of useful ways. What I want to discuss, however, is what we should make of this general class of work and how we can we make it better.

Let me note that this sort of work is multiplying. For example, I have been involved in a long line of work regarding the reliability of structural inferences based on long-run identifying restrictions in vector autoregressions (Faust and Leeper, 1997). At a presentation of my work with Eric Leeper in the early 1990s, Bob King asked why I did not assess the practical importance of the points using a DSGE model. I did not see the full merits of this at that time, but Erceg, Guerrieri, and Gust (2005) and Chari, Kehoe, and McGrattan (2008) have now taken up this suggestion (illustrating far more points than raised in the Faust-Leeper work) and considerably advanced the debate.

I would go so far as to argue that this sort of analysis should be considered a necessary component of best practice. That is, if anyone proposes a macroeconomic claim or advocates an econometric technique that is well defined in the new class of DSGE models, assessing the merits of the claim in the DSGE context should be mandatory. If it is coherent to apply the idea in the rat, we should do so before advocating its use in humans.

The work I am advocating cannot, however, be seen as part of some necessary or sufficient conditions for drawing *reliable conclusions* about reality. The mere fact that a particular claim is warranted in the DSGE model is neither necessary nor sufficient for the claim to be useful in practice. Similarly, the mere fact that a drug does not kill the rat is neither necessary nor sufficient for the drug's safety in humans. Just as judgment is required to draw lessons from animal studies, judgment will be required to draw lessons from

[1] Of course, many important issues remain in assessing this counterfactual, but these issues will not be important in this discussion.

[2] I add one important conceptual step in the discussion below.

DSGE studies. I believe that the results can be valuable nonetheless. In the remainder of the discussion, I highlight three points that, I believe, can make work in this spirit much more useful.

DOING IT BETTER

Don't Confuse the Rat with the Human

In animal studies, there is very rarely any confusion about when the authors are talking about the rats and when they are talking about the humans. The core of the research paper rigorously assesses some feature of the toxicology in rats and is clearly about the rat. Whatever one believes about the usefulness of the rat model, the point of the body of the paper is to support claims about the rat. This portion of the paper can be rigorously assessed without getting into unresolved issues about the ultimate adequacy of the rat model.

After settling issues about the rat, there is an active discussion about how the rat model results should be extrapolated to the human context.[3] This process is illustrated in the conclusions of a joint working group of the U.S. Environmental Protection Agency and Health Canada regarding the human relevance of animal studies of tumor formation (Cohen et al., 2004). They summarized their proposed framework for policy in the following four steps:

(i) Is the weight of evidence sufficient to establish the mode of action (MOA) in animals?

(ii) Are key events in the animal MOA plausible in humans?

(iii) Taking into account kinetic and dynamic factors, is the animal MOA plausible in humans?

(iv) Conclusion: Statement of confidence, analysis, and implications. (p. 182)

In the first step, we clarify the result in the model. The remaining steps involve asking serious questions about whether the transmission mechanisms in the model—to borrow a monetary policy term—plausibly operate similarly in the relevant reality.

In contrast, it is customary in macroeconomics to discuss quantities computed in the context of a DSGE model in a way that leaves it ambiguous at best whether the authors are advocating (or hoping) that we take them as statements about reality. I suspect that researchers arrive at the practice of treating statements about the model and reality as more-or-less equivalent under the rubric of "taking the model seriously." This seems to presume that the best way to take the model seriously is to take it literally. In toxicology, there is no doubt that policymakers take animal models seriously, but this never seems to require equating rats and humans. In my view, we should not confuse rats and humans; neither should we confuse DSGE models and reality.

Conceptual Clarity Before Computation

Broadly speaking, the point of the Cahn–Saint-Guilhem paper is to compare and contrast the behavior of two measures of potential output using a computational exercise on a DSGE model. Because it is so conceptually simple to implement computational experiments of the sort described above, it is very tempting to jump straight to the computer. I think work of this type would be better clarified by starting with careful conceptual analysis of the measures *before* computation. We can clearly lay out the expected differences and then many aspects of the computational work become exercises in measuring the empirical magnitude of effects that have been clearly defined.

I think this is particularly important in the macro profession where we seem to have a penchant for reusing labels for concepts that are quite distinct. "PFA potential" and "FPO potential" illustrate this point. "PFA potential" is meant to measure the level of output that would be attained if the current capital stock were used at some notion of "full capacity." "FPO potential" is, roughly speaking, the level of output that would be obtained if inputs were used *efficiently* as opposed to *fully*. It is clear that these two concepts of potential need not even be closely correlated. In any model in which the *efficient* level of, say,

[3] See Faust (2009) for a more complete discussion of this issue.

labor fluctuates considerably around the *full employment* level of labor, the two measures may be quite different. Clearly laying out the conceptual differences can be an incredibly enlightening step in what ultimately becomes a computational exercise.

One minor critique of the Cahn–Saint-Guilhem paper in this regard is that the work refers to FPO potential as simply the DSGE measure. There are many concepts of "potential" that might be useful for different questions in a DSGE model, and indeed we can discuss many versions of FPO potential, depending on how we implement the counterfactual regarding "if prices were flexible." Specific labels and careful analysis of the associated concepts can be very helpful.

Used properly, the sort of computational exercises with DSGE models that I am advocating can be an important tool for clarifying important conceptual issues. It may, at times, be tempting to simply substitute the relatively straightforward computational step for the sometimes painful step of careful conceptual analysis. Giving in to this temptation would be to miss an important opportunity.

Better Lab Technique

While the computational exercises I am advocating are conceptually straightforward, there are myriad subtle issues that fall under the umbrella of "lab technique." The new DSGE models are complicated and not fully understood. The Bayesian techniques being developed to analyze these models are also complicated and not fully understood. What we know from experience to date with DSGE models, and with similar tools applied in other areas, is that we can very easily create misleading results. For example, Sims (2003) has discussed such issues at length.

Much of the profession has long experience with the use of frequentist statistics and has become familiar with the myriad ways that one might inadvertently mislead. We need to be mindful of the fact that the profession is very new at assessing the adequacy of the new DSGE models using Bayesian techniques.

John Geweke (2005, 2007) has been at the forefront in developing flexible Bayesian tools for assessing model adequacy in the context of models that are known to be incomplete or imperfect descriptions of the target of the modeling. Abhishek Gupta (a Johns Hopkins graduate student) and I have recently been exploring these methods as they apply to DSGE models intended for policy analysis (Faust 2008, 2009; Faust and Gupta, 2009; and Gupta, 2009). I present just a flavor of one result with possible bearing on the topic of the Cahn–Saint-Guilhem analysis. The example is from Faust (2009), which reports results for the RAMSES model, a version of which is used by the Swedish Riksbank in its policy process.[4]

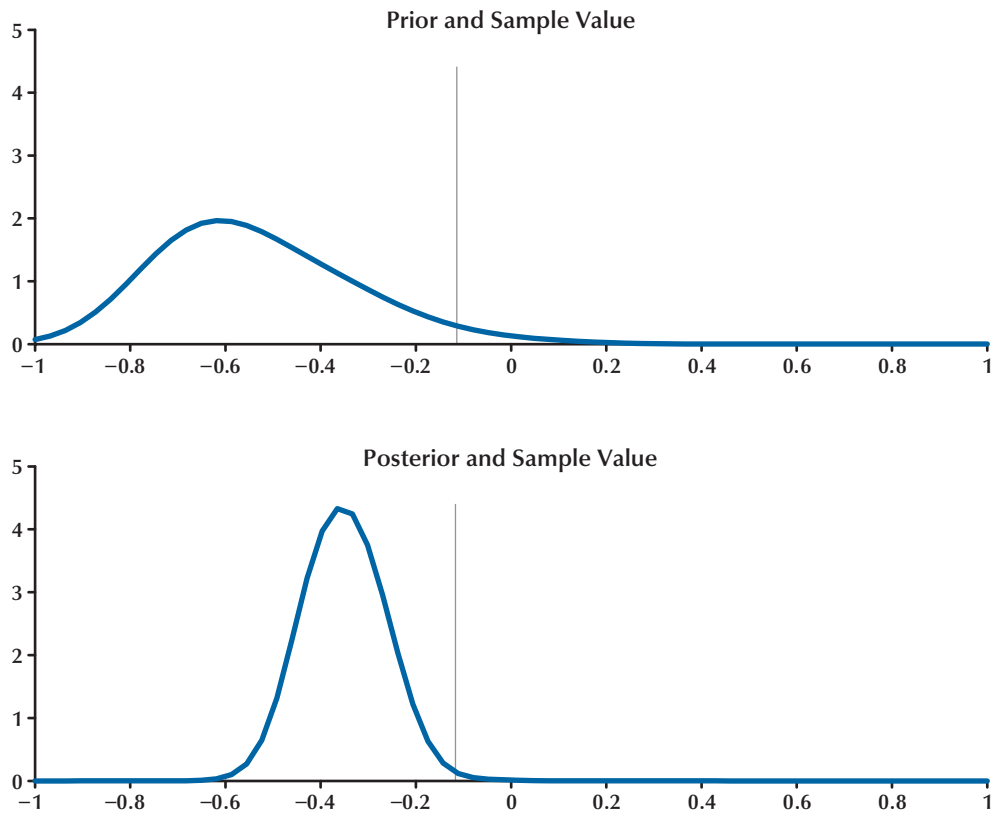
The simplest form of the idea is to take some feature of the data that is well defined outside the context of any particular macroeconomic model and about which we may have some prior beliefs. In the simplest form, we simply check whether the formal prior (which is largely arbitrary in current DSGE work) corresponds to our actual prior regarding this feature. Further, we check how both the formal prior and posterior compare with the data. A somewhat subtler version of this analysis instead considers prior and posterior predictive results for these features of interest.

As an example, consider the correlation between consumption growth and the short-term, nominally risk-free interest rate. Much evidence suggests that there is not a strong relation between short-term fluctuations in short-term rates and consumption growth. The upper panel of Figure 1 shows this marginal prior density implied by the prior over the structural parameters used in estimating RAMSES. The prior puts almost all mass on a fairly strong negative correlation, with the mode larger in magnitude than –0.5. The vertical line gives the value on the estimation sample of approximately zero. In short, the prior used in the analysis strongly favors the mechanism that higher interest rates raise saving and lower consumption. In the posterior (bottom panel), the mass is moved toward a negative correlation that is a bit smaller in magnitude, but the sample value is actually farther into the tail than it was in the prior.

[4] The developers of RAMSES (Riksbank Aggregated Macromodel for Studies of the Economy of Sweden) were exceedingly generous in helping me conduct this work.

Figure 1

Prior and Posterior Densities



NOTE: The figure shows the prior (upper panel) and posterior (lower panel) densities along with the sample value for the contemporaneous correlation between the short-term interest rate and quarterly consumption growth in a version of the RAMSES model.

SOURCE: Author's calculations using computer code provided by Riksbank.

This result and related ones in the work cited above convince me that current DSGE models continue to have difficulty matching basic patterns in consumption and investment as mediated by the interest rate. If we were to use this model in policy, we might want to ask whether this is one feature—like differences between rats and humans—that we should explicitly adjust for in moving from model results to reality.

Of course, the forces driving short-run fluctuations in consumption are at the very center of the distinction between PFA potential and FPO potential. These results and others like them convince me that while the DSGE model provides an interesting lab, there is good reason to question how literal we should be in extrapolating these results to the real economy.

A more general lesson is that the methods just sketched can be applied to any data feature, including statistics like those reported, for example, in Table 4 in the Cahn–Saint-Guilhem paper. These techniques allow one to coherently take estimation and model uncertainty into account and to evaluate the importance of arbitrary aspects in the formal prior. I strongly urge the authors to move in this direction.

CONCLUSION

I commend the St. Louis Fed for holding a conference on this issue that is vital to the monetary policymaking process, and I commend Christophe and Arthur for their interesting work illuminating how two competing measures of potential output behave in the context of modern DSGE models. This line of work is extremely important. I have made three suggestions that I believe would improve any work of this type. I hope that these suggestions contribute to making work of this sort even more influential.

REFERENCES

Cahn, Christophe and Saint-Guilhem, Arthur. "Issues on Potential Growth Measurement and Comparison: How *Structural* Is the Production Function Approach?" Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4) pp. 221-40.

Chari, V.V.; Kehoe, Patrick J. and McGrattan, Ellen R. "Are Structural VARs with Long-Run Restrictions Useful in Developing Business Cycle Theory?" NBER Working Paper No. 14430, National Bureau of Economic Research, October 2008; www.nber.org/papers/w14430.pdf.

Cohen, Samuel M.; Klaunig, James; Meek, M. Elizabeth; Hill, Richard N.; Pastoor, Timothy; Lehman-McKeeman, Lois; Bucher, John; Longfellow, David G.; Seed, Jennifer; Dellarco, Vicki; Fenner-Crisp, Penelope and Patton, Dorothy. "Evaluating the Human Relevance of Chemically Induced Animal Tumors." *Toxicological Sciences*, April 2004, *78*(2), pp. 181-86.

Erceg, Christopher J.; Guerrieri, Luca and Gust, Christopher. "Can Long-Run Restrictions Identify Technology Shocks?" *Journal of the European Economic Association*, December 2005, *3*(6), pp. 1237-78.

Faust, Jon. "DSGE Models in a Second-Best World of Policy Analysis." Unpublished manuscript, Johns Hopkins University, 2008; http://e105.org/faustj/download/opolAll.pdf.

Faust, Jon. "The New Macro Models: Washing Our Hands and Watching for Icebergs." *Economic Review*, March 23, 2009, *1*, pp. 45-68.

Faust, Jon and Gupta, Abhishek. "Bayesian Evaluation of Incomplete DSGE Models." Unpublished manuscript, Johns Hopkins University, 2009.

Faust, Jon and Leeper, Eric M. "When Do Long-Run Identifying Restrictions Give Reliable Results?" *Journal Business and Economic Statistics*, July 1997, *15*(3), pp. 345-53.

Geweke, John. *Contemporary Bayesian Econometrics and Statistics*. Hoboken, NJ: Wiley, 2005.

Geweke, John. "Bayesian Model Comparison and Validation." Unpublished manuscript, University of Iowa, 2007; www.aeaweb.org/annual_mtg_papers/2007/0105_0800_0403.pdf.

Gupta, Abhishek. "A Forecasting Metric for DSGE Models." Unpublished manuscript, Johns Hopkins University, 2009.

Sims, Chris. "Probability Models for Monetary Policy Decisions." Unpublished manuscript, Princeton University, 2003.

Parsing Shocks: Real-Time Revisions to Gap and Growth Projections for Canada

Russell Barnett, Sharon Kozicki, and Christopher Petrinec

The output gap—the deviation of output from potential output—has played an important role in the conduct of monetary policy in Canada. This paper reviews the Bank of Canada's definition of potential output, as well as the use of the output gap in monetary policy. Using a real-time staff economic projection dataset from 1994 through 2005, a period during which the staff used the Quarterly Projection Model to construct economic projections, the authors investigate the relationship between shocks (data revisions or real-time projection errors) and revisions to projections of key macroeconomic variables. Of particular interest are the interactions between shocks to real gross domestic product (GDP) and inflation and revisions to the level of potential output, potential growth, the output gap, and real GDP growth. (JEL C53, E32, E37, E52, E58)

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), 247-65.

Potential output is an important economic concept underlying the design of sustainable economic policies and decisionmaking in forward-looking environments. Stabilization policy is designed to minimize economic variation around potential output. Estimates of potential output may be used to obtain cyclically adjusted estimates of fiscal budget balances; projections of potential output may indicate trend demand for use in investment planning or trend tax revenues for use in fiscal planning; and potential output provides a measure of production capacity for assessing wage or inflation pressures.

Although potential output is an important economic concept, it is not observable. The Bank of Canada defines "potential output" as the sustainable level of goods and services that the economy can produce without adding to or subtracting from inflationary pressures. This definition is intrinsic to the methodology used by the Bank of Canada to construct historical estimates of potential output. In addition to using a production function to guide estimation of long-run trends influencing the supply side of the economy, the procedure incorporates information on the demand side that relates inflationary and disinflationary pressures to, respectively, situations where output exceeds and falls short of potential output.

Potential output and the "output gap," defined as the deviation of output from potential output, play central roles in monetary policy decisionmaking and communications at the Bank of Canada. Macklem (2002) describes the information and analysis presented to the Bank's Governing Council in the two to three weeks preceding a fixed announcement date.[1] As described in that document, the output gap—both its level and rate of change—is the central aggregate-demand

[1] In late 2000, the Bank of Canada adopted a system of eight preannounced dates per year when it may adjust its policy rate—the target for the overnight rate of interest. The Bank retains the option of taking action between fixed dates in extraordinary circumstances.

Russell Barnett was principal researcher at the time of preparation of this article, Sharon Kozicki is a deputy chief, and Christopher Petrinec is a research assistant at the Bank of Canada.

© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, the regional Federal Reserve Banks, or the Bank of Canada. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

link between the policy actions and inflation responses.[2]

In addition to being central to policy deliberations, the output gap has played a critical role in Bank of Canada communications. The concept of the output gap is simple to explain and understand. It has been used effectively to simultaneously provide a concise and intuitive view of the current state of the economy and inflationary pressures. It also provides a point of reference in relation to current policy actions and helps align the Bank's current thoughts on the economy with those held by the public.

Use of the output gap as a key communications device with the public is clearly seen in *Monetary Policy Reports* (*MPRs*) and speeches by governors and deputy governors of the Bank. The Bank of Canada began publishing *MPRs* semiannually in May 1995 (with two additional *Monetary Policy Report Updates* per year starting in 2000), and the output gap has been prominent in the reports from the beginning.[3] Indeed, a Technical Box appears in the first *MPR* regarding the strategy used by the Bank to estimate potential output.[4] Not only is the Bank's estimate of the output gap referenced in the text of the *MPR* as a source of inflationary (or disinflationary) pressure in the economy, but the estimates of recent history of the output gap up to the current quarter are also charted.

Governors and deputy governors have extensively used the output gap to explain to the general public how the monetary policy framework works. Common elements across these speeches include discussions on how potential output is estimated, how it is used to construct the output gap, and how the output gap affects monetary policy decisions. These discussions are nontechnical to enhance understanding by noneconomists. For instance, when discussing the factors affecting potential output in a speech to the Standing Senate Committee on Banking, Trade and Commerce in 2001, Governor David Dodge stated:

> [T]he level of potential rises over time as more workers join the labour force; businesses increase their investments in new technology, machinery and equipment; policy measures are taken to make product and labour markets more flexible; and all of us become more efficient and productive in what we do.

One important challenge associated with the use of potential output and the output gap as tools for communication of monetary policy decisions is that they cannot be directly observed and must be estimated. Moreover, estimates are prone to revision as historical data are revised and new information becomes available. Consequently, the Bank has directly addressed uncertainty surrounding estimates of the output gap and the drivers behind revisions in policy communications. A discussion of the implications of uncertainty for the conduct of monetary policy appeared in the May 1999 *MPR* (Bank of Canada, 1999, p. 26):

> [P]oint estimates of the level of potential output and of the output gap should be viewed cautiously. This has particular significance when the output gap is believed to be narrow and when inflation expectations are adequately anchored. In this situation, to keep inflation in the target range, policy-makers may have more success by placing greater weight on the economy's inflation performance relative to expectations and less on the point estimate of the output gap. At about the same time, the Bank started providing standard error bands around recent estimates of the output gap.[5]

---

[2] The important role of the output gap as a guide to monetary policymakers, over and above that of growth, was expressed by Governor Thiessen (1997):

> Some people apparently assume that it is the *speed* at which the economy is growing that determines whether inflationary pressures will increase or decrease. While the rate of the growth is not irrelevant, what really matters is the *level* of economic activity relative to the production capacity of the economy—in other words…the output gap in the economy. The size of the output gap, interacting with inflation expectations, is the principal force behind increased or decreased inflationary pressure.

[3] By contrast, incorporation of Governing Council projections has been more recent, with projections of core inflation first appearing in the April 2003 *MPR* and projections of gross domestic product (GDP) growth first appearing in the July 2005 *MPR*.

[4] The material in this box (May 1995) gives readers an idea of how the output gap is constructed without being overly technical. Publishing such statistics and the methods underlying their estimation has contributed importantly to monetary policy transparency in Canada.

[5] Standard error bands were provided around recent estimates from 1998 to 2007.

Revisions to historical estimates of potential output and the output gap also have been explicitly discussed in *MPR*s.[6] The discussions relate the revisions to recent developments in wage and price inflation and revised assessments of trends in labor input and labor productivity. Overall, transparency in the construction of the output gap, in understanding sources of revisions to past estimates of the output gap, and in uncertainty around the output gap has contributed to the effectiveness of the output gap as a key communications tool for enhancing understanding of the monetary policy process and of policy decisions in real time.

Implicit in the policy use of potential output and the output gap has been an effective strategy for managing volatility in estimates of the output gap. In particular, given the central role of potential output and the output gap in monetary policy, volatility in time series of the output gap or in revisions to estimates of the output gap can hinder the effectiveness of monetary policy communications, and therefore of monetary policy itself.

The next section reviews the methodology used by the Bank of Canada to estimate potential output and the output gap in Canada. While the methodology was designed to be consistent with the economic structure of the model used by Bank of Canada staff to construct projections, the Quarterly Projection Model (QPM), the procedure is designed to also incorporate information outside the scope of the model, such as demographics and structural details related to the labor market.[7] Features designed to contain end-of-sample revisions to estimates in response to updates of underlying economic data and to the availability of additional observations are discussed. This paper examines the extent to which such concerns were addressed by the methodologies developed to estimate and project potential output and the output gap in real time.

We next describe a dataset on real-time revisions to economic data and projections that has been constructed from a historical database of real-time economic projections made by Bank of Canada staff. The properties of these real-time revisions are explored in the subsequent text section. While the main focus of the analysis is the parsing of economic shocks into revisions to projections of (i) the level of potential output, (ii) the output gap, (iii) real GDP growth, and (iv) potential growth, the response of projections of inflation and short-term interest rates to shocks is also examined.

POTENTIAL OUTPUT IN CANADA

This section describes the techniques used by Bank of Canada staff to estimate historical values and project future values of potential output in Canada. In real time, Bank staff make ongoing marginal changes to the estimation methodology. Consequently, the description in this section should be taken only as broadly indicative of the procedures followed and the inputs to the estimation exercise.

A unifying assumption underlying both historical estimates and projections of potential output is that aggregate production can be represented by a Cobb-Douglas production function:

$$Y = (TFP)N^{a}K^{(1-a)}, \tag{1}$$

where Y is output, N is labor input, K is the aggregate capital stock, TFP is the level of total factor productivity, and a is the labor-output elasticity (or labor's share of income). This production function also was used in the now-discontinued model QPM to describe the supply side of the Canadian economy.

The next subsection describes the process by which historical estimates of potential output were estimated, while the following section focuses on assumptions underlying projections of potential output.

Historical Estimates of Potential Output

The methodology used to estimate potential output was heavily influenced by the requirements

[6] See, for instance, Technical Box 3 in Bank of Canada (2000).

[7] The QPM was used for economic projections between September 1993 and December 2005. Although there have been marginal changes in the procedure used to estimate the output gap over time, at the time of writing, the Bank continued to use basically the same methodology to generate its "conventional" estimate of the output gap in Canada.

of the monetary policy framework in which it was to be used. Thus, it was judged that the methodology should be consistent both with the QPM and the requirements associated with using the model to prepare economic projections. In this context, Butler (1996) notes that the following properties were judged to be of prime concern: consistency with the economic model (QPM); the ability to incorporate judgment in a flexible manner; the ability to both reduce and quantify uncertainty about the current level of potential output; and robustness to a variety of specifications of the trend component. In addition, given concerns about the feasibility and efficiency of estimates of potential output based solely on a model of the supply side of the economy, use of information from a variety of sources to better disentangle supply and demand shocks was deemed desirable.

With these guiding principles in mind, in the 1990s researchers at the Bank of Canada developed a new methodology to estimate potential output based on a multivariate filter that incorporates economic structure, as well as econometric techniques designed to isolate particular aspects of the data.[8] The main innovation was the development of a filter, known as the extended multivariate filter (EMVF), that solves a minimization problem similar to that underlying the Hodrick-Prescott (HP) filter (Hodrick and Prescott, 1997), but the EMVF also incorporates information on economic structure and includes modifications to penalize large revisions and excess sensitivity to observations near the end of the sample. For a variable or vector of variables, x, the general filter estimates the trend(s), x^*, as follows:

$$x^* = \max_{\hat{x}} \left\{ \left\{ -\left(x-\hat{x}\right)' W_x \left(x-\hat{x}\right) - \lambda \hat{x}' D' D \hat{x} \right\} + \left\{ -\varepsilon' W_\varepsilon \varepsilon - \left(s-\hat{x}\right)' W_s \left(s-\hat{x}\right) \right\} + \left\{ -\hat{x}' P' W_g P \hat{x} - \left(x_{pr}^* - \hat{x}\right)' W_{pr} \left(x_{pr}^* - \hat{x}\right) \right\} \right\} \tag{2}$$

This filter nests the HP filter, which is clearly evident for univariate x, by setting W_ε, W_s, W_{pr}, and W_g to zero, leaving only

$$\left\{ -\left(x-\hat{x}\right)' W_x \left(x-\hat{x}\right) - \lambda \hat{x}' D' D \hat{x} \right\}.$$

Information on economic structure and judgment can be introduced through the two terms

$$\left\{ -\varepsilon' W_\varepsilon \varepsilon - \left(s-\hat{x}\right)' W_s \left(s-\hat{x}\right) \right\}.$$

The term $\varepsilon' W_\varepsilon \varepsilon$ is the main channel through which information on the demand side of the economy may be introduced to assist in better separating demand shocks and supply shocks. In general, ε represents residuals from key economic relationships that depend on $\hat{x}$. For instance, if the unobserved trend to be estimated is the non-accelerating inflation rate of unemployment (NAIRU), ε may contain residuals from a Phillips curve that relate inflation developments to deviations of the unemployment rate from the NAIRU. In this sense, residuals may be interpreted as deviations from a structural economic relationship, perhaps drawing on cyclical economic relationships in the QPM. With this term in the filter the estimate of the trend may be shifted to reduce such deviations from the embedded economic theory.

Additional external structural information on trends may be introduced through the term $(s-\hat{x})' W_s (s-\hat{x})$. In this expression, s generally represents an estimate of the trend based on information outside the general scope of the model. For instance, in the case of the trend participation rate, s may be based on external analysis including information on demographics and otherwise informed judgment.

Finally, the last two terms,

$$\left\{ -\hat{x}' P' W_g P \hat{x} - \left(x_{pr}^* - \hat{x}\right)' W_{pr} \left(x_{pr}^* - \hat{x}\right) \right\},$$

provide a means to limit revisions to trend estimates. In general, procedures such as the EMVF are subject to one-sided filtering asymmetries at the ends of the sample. Although the filter is a symmetric two-sided weighted moving average within the sample period, near the end (and begin-

[8] See the discussion in Laxton and Tetlow (1992), Butler (1996), and St-Amant and van Norden (1997).

ning) of the sample, filter weights become one-sided. Intuitively, weights that would have been assigned to future observations if they were available are redistributed across recent observations. As a consequence, trend estimates near the end of the sample place large weights on recent data and tend to be revised considerably as additional observations become available.[9] The term $\hat{x}'P'W_gP\hat{x}$ penalizes large end-of-sample changes in the trend estimates and reduces the importance of the last few observations for the end-of-sample estimate of the trend. The term $(x^*_{pr} - \hat{x})'W_{pr}(x^*_{pr} - \hat{x})$ penalizes revisions to trend estimates between two successive projection exercises attributable to any source. In the absence of such a penalty, trend estimates could be revised more than is judged desirable due to (i) revisions to historical data, (ii) the availability of data for an additional quarter, or (iii) changes to external information or judgment as summarized by s.[10]

In many ways, the methodology of the EMVF was at the leading edge of research contributions in this area. For instance, although the methodology tends to be applied to estimate the trend in a single trending variable at a time, the theory is sufficiently general to include joint estimation of multiple trends, including situations with common trend restrictions. Stock and Watson (1988) developed a common trends representation for a cointegrated system, and state-space models as outlined in Harvey (1989) could also accommodate common trend restrictions. However, within the context of filters such as the HP filter, the band-pass filter of Baxter and King (1999), or the exponential-smoothing filter used by King and Rebelo (1993), imposition of common trend restrictions was not explored elsewhere in the academic literature until Kozicki (1999).

Another interesting aspect of the EMVF is that the methodology proposed approaches to reduce the importance of the one-sided filtering problem well before it was addressed elsewhere in the literature. Orphanides and van Norden (2002) drew attention to the result that many estimation methodologies yield large revisions to real-time end-of-sample estimates of the output gap. One potential approach to mitigating the one-sided filtering problem was proposed by Mise, Kim, and Newbold (2005).

As noted, an important characteristic of the EMVF is its ability to incorporate, within a filtering environment designed to extract fluctuations of targeted frequencies, information drawn from structural economic relationships, information from data sources external to the QPM, and judgment. The next few paragraphs provide details on the economic structure incorporated in the EMVF and the mechanism by which demographic information and structural features of the Canadian labor market could influence estimates of potential output.

Estimates of potential output are based on the Cobb-Douglas production function in equation (1). Recognizing that for the specification in equation (1), the marginal product of labor is $\partial Y/\partial N = aY/N$, the logarithm of output can be represented as

$$y = n + \mu - a, \tag{3}$$

where each term is expressed in logarithms and n is labor input, μ is the marginal product of labor, and $\alpha = \log(a)$ is the labor-output elasticity (also labor's share of income). The decision to use μ in constructions of historical estimates of potential output rather than data on the capital stock was motivated by concerns about the lack of timely (or quarterly) data and measurement problems. To construct log potential output, y^*, trends in log employment, n^*, the log marginal product of labor, μ^*, and the log labor share of income, α^*, are estimated separately and then summed.

One component of log potential output, trend log employment, n^*, is estimated using another decomposition:

$$n^* = Pop + p^* + \log\left(1 - u^*\right), \tag{4}$$

[9] Orphanides and van Norden (2002) show that revisions associated with the availability of additional data tend to dominate those related to revisions to historical data.

[10] An alternative possibility that was not explored was to penalize revisions of deviations from the trend rather than just revisions to the trend, by replacing the last term with the following:

$$((x - \hat{x}) - (x_{pr} - x^*_{pr}))'W_{pr}((x - \hat{x}) - (x_{pr} - x^*_{pr})).$$

In the absence of data revisions $x_{pr} = x$, all revisions to deviations would be due to revisions to the trend and both alternatives would yield the same results.

where *Pop* is the logarithm of the working-age population, p is the logarithm of the participation rate, and u^* is the NAIRU.[11] As for aggregate output, trend employment is constructed as the sum of the estimated trends of each component. The trend participation rate, p^*, is estimated with the EMVF using an external estimate of the trend participation rate for s, setting W_ε, W_{pr}, and W_ε to zero. Around the time of Butler's (1996) writing, the smoothness parameter λ was set to a very high number ($\lambda = 16000$) to obtain a very smooth estimate of the trend participation rate. However, the value of this parameter has been adjusted considerably over time and more recently has been set to $\lambda = 1600$, a value typically used to exclude fluctuations of "typical" business cycle frequencies from trend estimates. The external estimate of the trend participation rate accounts for demographic developments, including, for instance, trends in the workforce participation rate of women and school employment rates.[12] The NAIRU is also estimated using the EMVF, with an external estimate of the trend unemployment rate based on the work of Côté and Hostland (1996) used for s, and residuals ε, obtained from a price-unemployment Phillips curve drawing on the work of Laxton, Rose, and Tetlow (1993). The external estimate of the trend unemployment rate incorporates information on structural features of Canadian labor markets, including the proportion of the labor force that is unionized and payroll taxes.

A second component of log potential output is the trend value of the log labor-output elasticity, a^*. This component is estimated as the smooth trend obtained by applying an HP filter with a large smoothing parameter ($\lambda = 10000$) to data on labor's share of income.

The third component of log potential output, the trend log marginal product of labor, μ^*, is also estimated by applying the EMVF. The real producer wage is used for s rather than an external estimate of the trend, and ε is the residual from an inflation/marginal product of labor relationship. The latter is motivated by the idea that the deviation of the marginal product of labor from its trend level can be interpreted as a factor utilization gap and, hence, provides an alternative index of excess demand pressures.

Projecting Potential Output

Projections of potential output are based on the Cobb-Douglas production function, equation (1), but are driven by consideration of supply-side features:

$$y^* = tfp^* + a^* n^* + \left(1 - a^*\right) k, \tag{5}$$

where lower-case letters indicate the logarithm of the respective capitalized notation and an asterisk denotes that a variable is set to its trend or equilibrium value.[13] Thus, projections of potential output are constructed with projections of tfp^*, a^*, n^*, and k.

The capital stock, k, is constructed from the cumulated projected investment flows given the actual capital stock at the start of the projection. The equilibrium labor-output elasticity, a^*, is set to a constant equal to the historical average labor share of income.

The typical assumption is that in the medium to long term, trend total factor productivity, tfp^*, will converge toward the level of productivity of the United States at the historical rate of convergence.[14] A short-run path for tfp^* links the historical estimate at the start of the projection to the medium-term path for tfp^*, with short-run behavior based on typical cyclical variation.

The equilibrium employment rate, n^*, is based on an analysis of population growth, labor force participation, and structural effects on the NAIRU (Bank of Canada, 1995). Analysis draws on information outside the scope of the QPM. For instance, labor force participation is related to demographic factors (Bank of Canada, 1996); population growth may be influenced by immi-

[11] Barnett (2007) provides recent estimates and projections of trend labor input using a cohort-based analysis that incorporates anticipated demographic changes. Barnett's analysis also accounts for trend movements in hours.

[12] See Technical Box 2 of Bank of Canada (1996).

[13] See the discussion in Butler (1996).

[14] Crawford (2002) discusses determinants of trends in labor productivity growth in Canada.

gration policy; and the NAIRU may be related to structural factors.[15] To a large extent, this series can be thought of as corresponding to an external structural estimate of s as used in the EMVF. Thus, projections are as if they are generated from an application of the EMVF with all weights other than W_s set to zero.

Although numerous studies—including Butler (1996), Guay and St-Amant (1996), St-Amant and van Norden (1997), and Rennison (2003)—have compared the properties of alternative approaches of historical estimates of the output gap across alternative estimation approaches, no similar studies exist to examine properties of projections of potential output or the output gap. This is one area to which the current study hopes to contribute.

MEASURING SHOCKS AND REVISIONS TO PROJECTIONS

The empirical analysis is designed to assess the sensitivity of economic projections to new information. If economic projections were "raw" outputs from application of the QPM, then our analysis would be merely recovering information about the structure of the QPM, which is available elsewhere.[16] However, in general, economic projections are influenced by judgment to account for features of the economy outside the scope of the economic model. In addition, the QPM is primarily a business cycle model, designed to project deviations of economic variables from their respective trend levels. Consequently, while potential output and other trends are constructed to be consistent with the economic structure of the QPM, evolution of these trends is modeled outside the QPM.

Real-Time QPM Projections and Data

The analysis uses real-time data from the Bank of Canada's staff economic projection database. Bank staff generate projections quarterly to inform the policy decisionmaking process. The projection data analyzed in this project were generated by the QPM. It is important to note that the projections in these data correspond to staff economic projections and may not be the same as projections implicitly underlying policy decisions, or, in later years, as published in the *MPR*, as such projections would correspond to the views of the Governing Council.

Analysis is limited to projection data for the period September 1993 through December 2005, the period during which the QPM was used by Bank staff producing projections. By limiting empirical analysis to data within this period, the likelihood of structural breaks in projections associated with large changes in the projection model is small. An additional advantage of this sample is that it falls entirely within the inflation-targeting regime in Canada, removing concerns about structural breaks associated with changes in policy regime.

The database includes a total of 50 vintages of data, one vintage for each quarterly projection exercise. As is standard in the real-time-data literature, the term "vintage" is used to refer to the dataset corresponding to the data from a specific projection. Vintages are described by the month and year when the projection was made. Projections were generated four times per year, once per quarter, in March, June, September, and December. For each vintage, the database contains the history of the conditioning data as available at the time of the projection, as well as the projections.

This database is used to construct measures of shocks and projection revisions. Both shocks and revisions are constructed as the difference between values of economic variables (either historical observations or the projection of a specific variable) for a given quarter as recorded in two successive vintages of data. The term "revision" is reserved to reflect a change in the projection of a variable, whereas the term "shock" is used to reflect the difference between a new or revised

[15] Poloz (1994) and Côté and Hostland (1996) discuss the effects of structural factors, such as demographics, unionization, and fiscal policies influencing unemployment insurance, the minimum wage, and payroll taxation, on the NAIRU. More information on demographic implications for labor force participation is provided by Ip (1998).

[16] A nontechnical description of the QPM is provided in Poloz, Rose, and Tetlow (1994). Detailed information on the QPM is provided in the trio of Bank of Canada Technical Reports by Black et al. (1994), Armstrong et al. (1995), and Coletti et al. (1996).

observation for a variable and its value (either an observation or a projection) as recorded in the previous vintage of data. For each economic variable, 2 sets of shocks series and 12 sets of revisions series are constructed.

The timing of the publication of data is critical to understanding the distinction between shocks and revisions. In general, data for a full quarter, t, are not published until the next quarter, $t + 1$. Thus, for instance, in the month when Bank of Canada staff were conducting a projection exercise, the values of variables recorded for the current quarter were "0-quarter-ahead" projections; values for the next quarter were "1-quarter-ahead" projections; and values for the prior quarter were published data. Letting x_t^v denote the value of variable x for quarter t as recorded in vintage v of the dataset, x_t^v denotes a $(t - v)$-quarter-ahead projection for $t \geq v$ and is treated as an observation of published data if $t < v$. The term "published" is somewhat of a misnomer and is more appropriate for data on inflation, real GDP, and interest rates, for instance, than for potential output, potential growth, or the output gap as the latter three concepts are not directly observed, nor are they measured or constructed by the statistical agency of Canada, Statistics Canada. As discussed earlier, values of these variables are estimated internally by Bank of Canada staff. Nevertheless, for notational convenience and to facilitate parsimonious exposition, language such as "observation," "data," and "published" is used synonymously in reference to all series according to the timing convention previously described.

The term "shock" is generally used to refer to marginal information from one vintage to the next provided by new observations on market interest rates, new or updated data produced by Statistics Canada, or new or updated historical estimates of potential output (and related series) constructed by the Bank of Canada. Two measures of shocks are examined:

$$shock1_t = x_t^{t+1} - x_t^t \tag{6}$$

is the difference between the published value of variable x for quarter t as available in quarter $t + 1$ (the first quarter it is published) and the 0-quarter-ahead (or contemporaneous) projection of variable x as made in t and recorded in vintage $v = t$. Thus, *shock1* is a projection error. The second measure of shocks captures the first quarterly update to the published data and is constructed as

$$shock2_t = x_{t-1}^{t+1} - x_{t-1}^t. \tag{7}$$

The term "revisions" is used to refer to changes in Bank of Canada staff projections of a variable between successive vintages. Twelve measures of revisions are examined with each corresponding to a different projection horizon,

$$revisionk_t = x_{t+k+1}^{t+1} - x_{t+k+1}^t, \tag{8}$$

where $k = 0,\ldots 11$.

The analysis in this article concentrates on shocks and revisions to nine variables as defined below:

- *EXCH*: the bilateral exchange rate between Canada and the United States, expressed as $US per $CDN;
- *GAP*: the output gap defined as the percent deviation of real GDP from potential real GDP (potential output);
- *GDP*: real GDP growth (an annualized quarterly growth rate);
- *GDPLEV*[17]: log-real GDP level, constructed as an index for a given quarter by taking *GDPLEV* for the prior quarter and adding (100/4) * log(1 + *GDP*) to it, with current vintage data for a given quarter early in the sample used to initiate the recursive construction;
- *POT*: potential output growth, calculated as the annualized one-period percent change in *POTLEV*;
- *POTLEV*: log potential output level, constructed as *GDPLEV* – *GAP*, an index;
- *INF*: CPI inflation (annualized quarterly growth rate);

[17] During the period of analysis, Statistics Canada rebased GDP several times. From 1994 to 1996, the base year used for real GDP calculations was 1986. The base year changed to 1992 from 1996 to July 2001. From July 2001 to May 2007, the base year used was 1997. However, as *GDPLEV* and *POTLEV* were constructed as indices, these rebasings would not affect the analysis of this study.

- *INFX*: core CPI inflation (annualized quarterly growth rate). The definition of core CPI has changed over our period of analysis. Before May 2001 the Bank of Canada used CPI excluding food, energy, and indirect taxes (CPIxFET) as the measure of core inflation. After May 2001 the Bank changed its official measure of core inflation to CPI excluding the eight most volatile components (CPIX), and;
- *R*90: a nominal 90-day short-term interest rate.

The information content of contemporaneous, $k = 0$, projections will differ across variables projected, implying that for some variables *shock1* will be much smaller than for others. In particular, projections are made in the third month of each quarter. However, the initial release of the national accounts is at the end of the second month or early in the third month of the quarter. Data in these national accounts releases, such as GDP, extend only through the prior quarter. For example, for the national accounts release in late August 2008 (the second month of Q3), the latest GDP observations are for 2008:Q2. However, some statistics are available in a more timely manner. For example, interest rate data are available in real time. Thus, by the third month in a quarter, two months of interest rate data are already available. Likewise, for some variables *shock2* will be much smaller (and in some cases zero) than for others, because some published data series, such as GDP, are revised in quarters after the initial release, while others, such as interest rates, are not.

PROPERTIES OF PROJECTION REVISIONS

New information becomes available in the period between projection exercises. This information takes many forms, including new or revised data published by statistical agencies, new observations from financial markets, as well as anecdotal information from surveys or the press, among others. For interest rates, inflation, and real GDP growth, the information in *shock1* reflects projection errors, whereas the information in *shock2* reflects revised data. By contrast, shocks to potential output, the output gap, and potential growth generally are a function of shocks to data (including, but not limited to, interest rates, inflation, and real GDP growth), updated judgment on the part of Bank of Canada staff, and updates to external structural information on trends. Revisions may reflect some or all of the varying types of new information. New observations of some published data directly enter into model projections, but other information may inform judgment and also be incorporated.

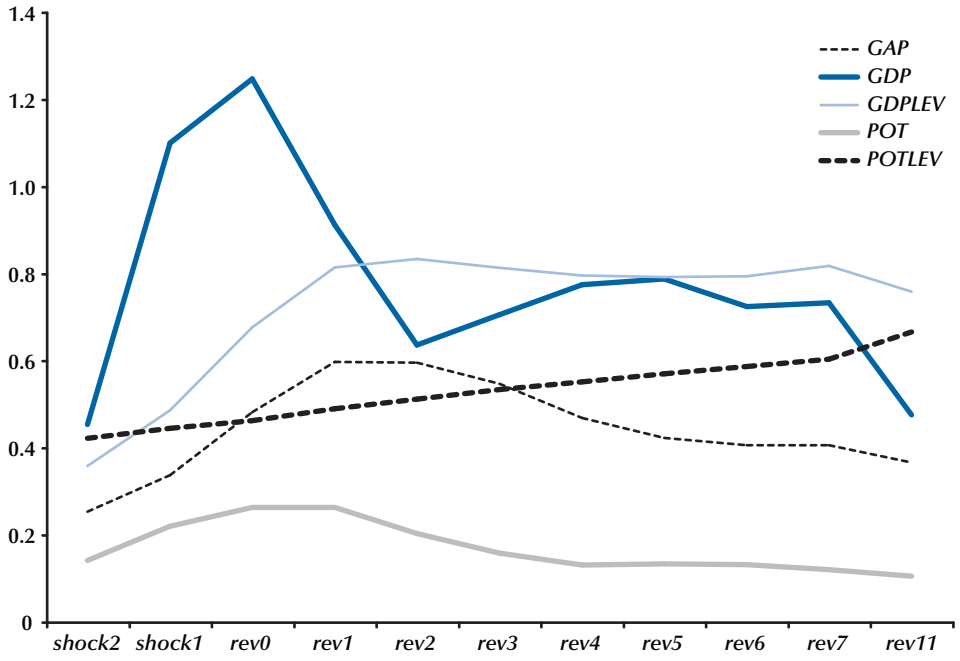
This section examines the properties of shocks and revisions. The analysis examines the relative size of revisions to projections of trends compared with revisions to projections of cyclical dynamics. Another issue of particular interest is the parsing of shocks to real GDP growth, interest rates, inflation, and exchange rates into permanent and transitory components that will, in turn, affect shocks and revisions of projections of potential output and the output gap.

Properties of Projection Revisions and Shocks

Figure 1 shows the standard deviations of shocks and revisions to *GAP*, *GDP*, *GDPLEV*, *POT*, and *POTLEV*. This figure shows that both shocks and revisions to potential output growth (*POT*) were small at all horizons. By contrast, projection errors (*shock1*) and near-term revisions to real GDP growth (*GDP*) tend to be considerably larger. Both results are consistent with what would generally be expected. By definition, potential is meant to capture low-frequency movements in output and is constructed to be smooth. Consequently, it would be surprising to see either volatile potential growth or frequent large revisions to potential growth. Real GDP growth, however, tends to be volatile. Thus, not surprisingly, revisions, particularly to current and one-quarter-ahead projections, can be sizable. Much of the volatility of both the underlying growth rate data and the revisions is likely related to the allocation and reallocation of inventory investment, imports, and exports across quarters. At longer horizons,

Figure 1

Standard Deviations



the standard deviation of *GAP* projection revisions remains quite large, and the standard deviation of revisions to projections of GDP growth are considerably larger than revisions to projections of potential growth. These observations suggest considerable persistence in business cycle propagation of economic shocks. Even at a 2- to 3-year horizon, real GDP growth does not consistently converge to potential output growth in projections.

Whereas shocks and revisions to potential growth are considerably smaller than revisions to GDP growth, the same is not true for the log levels of GDP (*GDPLEV*) and potential output (*POTLEV*). For these variables, the standard deviations of shocks are essentially the same. As expected, *GDPLEV* revisions tend to be larger than *POTLEV* revisions, but not by nearly as much as was the case for their growth rates (*GDP* and *POT*, respectively). In fact, at the longest horizon, $k = 11$, the magnitudes of revisions to the levels are, on average, essentially the same.

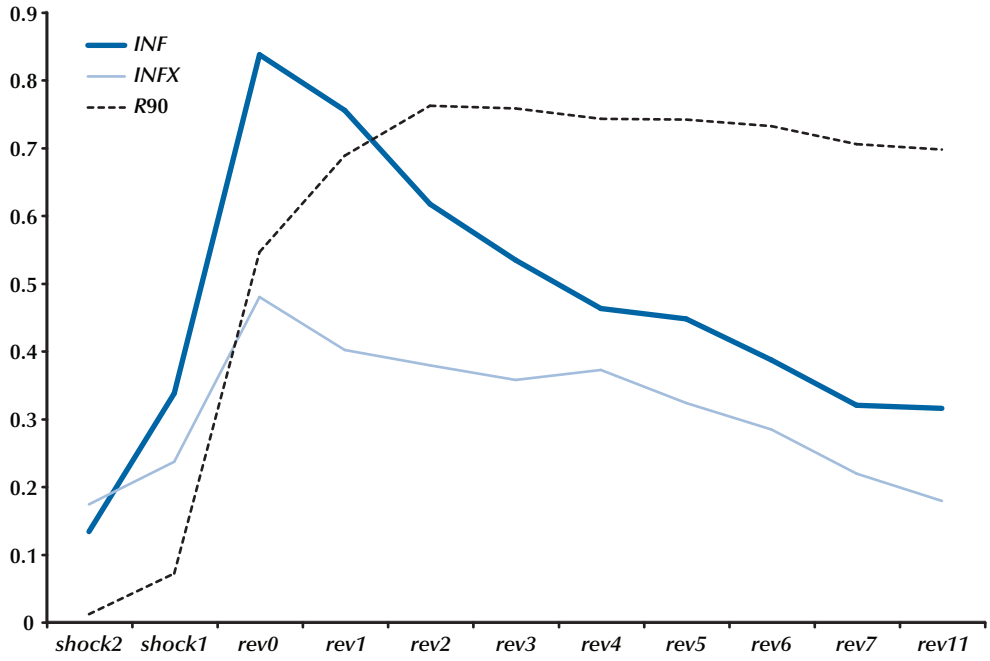
Figure 2 shows the standard deviations of shocks and revisions to *INF*, *INFX*, and *R*90. Shocks to all variables are quite small. As noted earlier, some monthly data are available for the contemporaneous quarter, likely explaining the larger differences in standard deviations of the projection error (*shock1*) relative to the first forecast revision (*rev0*).[18] Revisions to near-term projections tend to be larger than those to longer-horizon projections for inflation. This may reflect the effects of endogenous policy designed to achieve the 2 percent target at a roughly 2-year horizon.

Very different properties are evident for the short-term interest rate (*R*90). Shocks to interest rates are generally small, owing to the fact that interest rate data are available daily in real time (so that much of the current-quarter information

[18] For *INF* and *INFX*, annual updates to seasonal adjustments to the data are the main source of nonzero values of *shock2*. The change in definition of *INFX* in May 2001 also leads to a nonzero value of *shock2* for this variable.

Figure 2

Standard Deviations



is already available at the time of the contemporaneous-quarter projections) and are not very volatile. Standard deviations of *revision0* are similar to those of inflation. However, as the forecast horizon increases, standard deviations of revisions to interest rates rise somewhat before leveling off, and in contrast to the results for inflation, they do not noticeably decline for longer forecast horizons.

Table 1 provides information on the persistence of projection revisions across forecast horizons.[19] Persistence should vary considerably across different economic variables. In general, revisions to trend levels should be expected to be permanent, while revisions to cyclical variables should be expected to dissipate. Each column of Table 1 provides correlations of shocks and revisions with *revision0* for a single variable. When *revision0* of *GAP* is revised, so are revisions to *GAP* projections at other horizons, although the correlation diminishes as the projection horizon increases. Potential growth revisions are also positively correlated but display a somewhat different pattern, with much lower correlation at near-term horizons. GDP growth revisions show strong near-term momentum, but negative correlations suggest near-term revisions tend to be partially reversed further out.

Correlations across horizons of revisions to projections of the three level variables, *GAP*, *GDPLEV*, and *POTLEV*, clearly reveal the differing persistence properties of trends and cycles. When the level of potential output is revised, it tends to be revised by nearly equal amounts at all projection horizons. By contrast, as noted previously, when the contemporaneous-quarter projection of the output gap is revised, subsequent projections are revised in the same direction, but by diminishing amounts as the projection horizon increases. By construction, *GDPLEV* is the sum

[19] Note that an alternative definition of persistence would examine the persistence of revisions by horizon across time.

Table 1

Correlations of Revisions Across Projection Horizons

Revision	Gap	GDP growth	Potential growth	CPI inflation	Core inflation	Short-term interest rate	GDP level	Potential level	CPI level	Core CPI level	Exchange rate
Shock2	0.30	0.34	0.26	0.04	–0.11	0.16	0.70	0.98	0.64	0.97	–0.06
Shock1	0.79	0.50	0.56	0.53	0.38	0.08	0.91	0.99	0.80	0.98	0.35
Rev0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Rev1	0.94	0.71	0.36	0.09	0.19	0.76	0.97	0.99	0.90	0.99	0.92
Rev2	0.87	0.09	0.31	–0.10	0.32	0.55	0.93	0.98	0.82	0.96	0.87
Rev3	0.81	–0.17	0.29	–0.19	0.23	0.44	0.89	0.98	0.71	0.93	0.79
Rev4	0.69	–0.46	0.26	–0.12	0.22	0.42	0.84	0.97	0.62	0.90	0.70
Rev5	0.53	–0.44	0.25	–0.01	0.17	0.38	0.77	0.96	0.55	0.87	0.65
Rev6	0.28	–0.48	0.34	0.06	0.10	0.33	0.69	0.95	0.50	0.84	0.62
Rev7	0.01	–0.43	0.29	0.13	0.02	0.25	0.61	0.95	0.48	0.82	0.60
Rev11	–0.25	–0.04	0.06	0.45	–0.10	0.09	0.55	0.93	0.53	0.77	0.53

Table 2

Correlations among Projection Errors (*shock1*)

	Gap	GDP growth	Potential growth	Potential log level	CPI inflation	Core inflation	Short-term interest rate	Exchange rate
Gap	1.00	0.62	–0.08	–0.25	0.14	0.05	–0.01	–0.01
GDP growth	0.62	1.00	0.32	0.29	0.01	–0.01	0.07	–0.01
Potential growth	–0.08	0.32	1.00	0.47	0.09	–0.14	–0.13	0.11
Potential log level	–0.25	0.29	0.47	1.00	–0.15	–0.09	0.01	0.07
CPI inflation	0.14	0.01	0.09	–0.15	1.00	0.63	–0.09	–0.17
Core inflation	0.05	–0.01	–0.14	–0.09	0.63	1.00	–0.14	–0.31
Short-term interest rate	–0.01	0.07	–0.13	0.01	–0.09	–0.14	1.00	–0.02
Exchange rate	–0.01	–0.01	0.11	0.07	–0.17	–0.31	–0.02	1.00

Table 3

Correlations among Data Revisions (*shock2*)

	Gap	GDP growth	Potential growth	Potential log level
Gap	1.00	–0.07	–0.35	–0.53
GDP growth	–0.07	1.00	0.46	0.31
Potential growth	–0.35	0.46	1.00	0.59
Potential log level	–0.53	0.31	0.59	1.00

of *POTLEV* and *GAP*, so it should not be surprising that persistence properties are intermediate to the two components. On average, about half of the contemporaneous projection revision is permanent, whereas the other half shrinks with longer forecast horizons.

This result is rather striking, as is the result (evident in Figure 1) that the standard deviation of shocks to the level of GDP is about the same as the standard deviation of shocks to the level of potential GDP. Moreover, the standard deviations of revisions to projections of the level of potential output are only somewhat smaller than the standard deviations of revisions to projections of the output gap (and are smaller for only three near-term forecasting horizons). Overall, these results suggest almost the same amount of uncertainty is associated with the level of potential as with the gap. Of course, all else equal, revisions to the level of potential output do not have policy implications, whereas revisions to the output gap do.

In the case of inflation, revisions to core inflation projections tend to have some, albeit low, persistence, whereas those to overall inflation do not. This result is consistent with the observation that near-term revisions to overall inflation are generally driven by information on the volatile components excluded from the core measures. By contrast, revisions to exchange rate projections are very persistent. The persistence of revisions to projections of the short-term interest rate is roughly similar to the persistence of revisions to the gap, perhaps indicating a link between the two. This possibility is explored in the next subsection.

Correlations among projection errors (*shock1*) are presented in Table 2. A few interesting results emerge. First, correlations among projection errors to GDP growth, core inflation, *R*90, and the exchange rate are very low. Second, the correlation between projection errors to GDP growth and the output gap is quite high. This result likely signals

Table 4

Regression Results: Responses of Revisions to Shocks

Dependent variable (*revisionk*)	*k*	*GDP shock1*	*GDP shock2*	*INFX shock1*	*R90 shock1*	*EXCH shock1*	$\bar{R}^2$
GAP	0	0.30***	0.17	–0.22	–2.13***	0.12	0.65
	1	0.31***	0.33**	–0.22	–2.69***	0.16	0.57
	2	0.28***	0.38**	–0.20	–2.66***	–0.03	0.51
	3	0.24***	0.34**	–0.25	–2.29**	–0.11	0.46
	4	0.18***	0.29**	–0.23	–1.65*	–0.14	0.38
	5	0.13**	0.23	–0.24	–1.13	–0.12	0.27
	6	0.09	0.15	–0.19	–0.80	–0.15	0.14
	7	0.03	0.11	–0.15	–0.78	–0.04	0.05
	11	–0.02	0.08	0.11	–1.57**	0.32	0.16
GDP	0	0.47***	0.73**	–1.24*	–7.61***	0.48	0.49
	1	0.07	0.66**	–0.36	–1.79	–0.05	0.14
	2	–0.16*	0.13	–0.04	0.29	–0.71	0.10
	3	–0.18*	–0.14	–0.27	2.08	–0.32	0.16
	4	–0.24**	–0.21	0.12	3.10*	–0.11	0.21
	5	–0.17	–0.26	0.00	2.41	0.12	0.14
	6	–0.16	–0.29	0.26	1.83	–0.01	0.14
	7	–0.24**	–0.13	0.20	0.49	0.50	0.17
	11	–0.04	–0.14	0.02	–0.23	0.21	0.05
POT	0	0.02	0.00	–0.02	0.34	0.08	0.02
	1	0.02	–0.01	–0.37*	0.53	–0.24	0.14
	2	–0.02	–0.05	–0.12	0.19	0.08	0.06
	3	–0.01	–0.00	–0.05	0.50	0.02	0.06
	4	0.00	0.00	0.03	0.40	0.01	0.05
	5	0.01	0.03	0.04	0.26	0.07	0.05
	6	0.01	0.05	0.06	0.42	0.08	0.12
	7	0.02	0.03	0.04	0.42	0.07	0.13
	11	–0.01	0.10***	–0.03	0.34*	–0.07	0.28
POTLEV	0	0.08	0.26	–0.17	–0.38	0.15	0.15
	1	0.08	0.26	–0.26	–0.25	0.09	0.14
	2	0.08	0.25	–0.29	–0.20	0.11	0.13
	3	0.08	0.25	–0.30	–0.08	0.11	0.12
	4	0.08	0.25	–0.29	0.01	0.12	0.11
	5	0.08	0.26	–0.28	0.08	0.13	0.11
	6	0.08	0.27	–0.27	0.18	0.15	0.11
	7	0.08	0.27	–0.26	0.28	0.17	0.12
	11	0.08	0.36	–0.28	0.61	0.16	0.13

*Significant at 10 percent; **significant at 5 percent; ***significant at 1 percent.

Table 4, cont'd

Regression Results: Responses of Revisions to Shocks

Dependent variable (*revisionk*)	*k*	*GDP shock1*	*GDP shock2*	*INFX shock1*	*R90 shock1*	*EXCH shock1*	$\bar{R}^2$
POTLEV	0		0.35**	–0.17	–0.41	0.14	0.12
	1		0.35**	–0.26	–0.28	0.09	0.12
	2		0.33*	–0.29	–0.24	0.10	0.10
	3		0.33*	–0.31	–0.11	0.11	0.10
	4		0.33*	–0.30	–0.02	0.11	0.09
	5		0.34*	–0.29	0.04	0.13	0.09
	6		0.35*	–0.28	0.14	0.15	0.09
	7		0.36*	–0.27	0.24	0.16	0.10
	11		0.44**	–0.29	0.57	0.16	0.12
INF	0	–0.01	0.22	0.78	–0.68	0.73	0.07
	1	0.07	0.63**	–0.01	–0.51	–0.02	0.18
	2	0.14	–0.03	0.56	–1.75	0.25	0.15
	3	0.18**	0.03	0.32	–1.00	0.32	0.18
	4	0.10	0.16	0.57*	–1.35	0.23	0.23
	5	0.14**	0.04	0.57**	–1.01	0.42	0.24
	6	0.10**	0.07	0.61***	–1.00	0.29	0.28
	7	0.11***	0.07	0.50***	–0.99*	0.33	0.37
	11	0.07	–0.06	0.25	0.40	0.38	0.10
INFX	0	–0.06	0.03	0.71**	–0.89	–0.09	0.18
	1	0.09	0.19	0.04	–0.69	–0.13	0.15
	2	0.09*	0.14	–0.03	–1.65**	–0.23	0.21
	3	0.10**	0.22*	0.14	–1.09	–0.21	0.28
	4	0.04	0.34***	0.34	–1.05	–0.14	0.30
	5	0.08*	0.20*	0.33*	–0.86	–0.06	0.30
	6	0.07*	0.15	0.33*	–0.36	–0.04	0.28
	7	0.06**	0.13*	0.20	–0.23	0.01	0.29
	11	0.03	–0.01	0.05	0.43	0.30*	0.15
R90	0	0.01	0.21	–0.04	0.30	–0.09	0.04
	1	0.16*	0.51**	–0.06	–1.57	–0.11	0.25
	2	0.28***	0.61***	0.04	–0.41	0.32	0.41
	3	0.28***	0.60***	–0.00	–0.03	0.46	0.43
	4	0.25**	0.48**	–0.20	0.81	0.44	0.35
	5	0.22**	0.34	–0.25	1.03	0.68	0.26
	6	0.17	0.28	–0.18	1.05	1.01*	0.22
	7	0.15	0.19	–0.01	1.09	1.11*	0.19
	11	0.08	0.02	0.25	0.52	1.19*	0.10

*Significant at 10 percent; **significant at 5 percent; ***significant at 1 percent.

that for a given level of potential output, higher than expected GDP data would raise both GDP growth and the gap. Similarly, the correlation between projection errors to CPI inflation and core inflation is high, consistent with the fact that CPI inflation is an aggregate that contains core inflation, so that errors in core inflation would also show up in CPI inflation. Correlations among data revisions (Table 3) are of the same sign as those among projection errors, although the former are generally stronger.

Trend versus Cycle: Projection Revisions in Response to Shocks

An important element of projection exercises is parsing shocks into permanent components that influence trends but do not have inflationary consequences, and transitory components that affect cyclical dynamics and generally affect inflationary pressures. The QPM was the primary tool used to map the implications of transitory structural shocks into economic projections. While judgment may have also entered into projections, particularly for understanding near-term economic variation, at medium to longer horizons, endogenously generated model dynamics would play a more dominant role. As noted earlier, the properties of the QPM are well documented. However, the implications of shocks for trend projections are less well understood.

In this section, the responses of projections of several main economic variables to shocks to *GDP*, *INFX*, *R*90, and *EXCH* are analyzed. To a certain extent, shocks to these variables might be considered exogenous, as they directly reveal new information from financial markets (in the case of interest rates and exchange rates) or as published by Statistics Canada. Revisions to potential output (and variables constructed using potential output) might be thought of as responses to this new information.[20] To assess the importance of these sources of new information, regressions of the following format were estimated:

$$\begin{aligned} revisionk_t &= c + \beta_{G1} GDPshock1_t \\ &+ \beta_{G2} GDPshock2_t + \beta_I INFXshock1_t \\ &+ \beta_R R90shock1_t + \beta_E EXCHshock1_t . \end{aligned} \tag{9}$$

Only one shock variable was included for inflation, the short-term interest rate, and the exchange rate, as these variables are essentially unrevised. Results are presented in Table 4.

The most important variable in terms of influencing projection revisions is *GDP*. Shocks to *GDP* tend to lead to revisions of the same sign to projections of the output gap, inflation, core inflation, the short-term interest rate, the level of potential output, and near-term projections of real GDP growth; and to revisions of the opposite sign to longer-term projections of real GDP growth. By contrast, there is no evidence that potential growth is responsive to these shocks. In terms of parsing *GDP* shocks, a fraction of these shocks (about 1/3) are mapped into permanent shocks that lead to parallel shifts of the level of potential without influencing the growth rate. The remainder of the *GDP* shocks are assessed as cyclical (transitory), with some persistence, and lead to revisions to gap projections at horizons out to five quarters, with the largest revisions being to *revision1* and *revision2* (about 2/3 of *GDP* shocks are mapped into *GAP* revisions for $k = (1,2)$). For positive shocks, near-term growth is revised upward and the output gap becomes larger. The additional inflationary pressures lead to tighter monetary policy, which is consistent with more rapid reductions in the size of the gap and therefore downward revisions to GDP growth, both of which are consistent with a closing of the gap after two years.

There are two noteworthy aspects to this parsing of *GDP* shocks into potential output and the output gap. First, parsing explicitly recognizes that not all shocks are transitory demand shocks. In the EMVF filter, the HP terms imply that estimates of potential output are informed by historical output data. Thus, shocks may lead to revised estimates of potential output for the last few observations of the historical data. The empirical results

[20] In examining the empirical results in the table, it is important to keep in mind that some shocks have smaller standard deviations than others. In particular, because interest rates tend to move gradually and two of three months of interest rate data are available for the contemporaneous quarter during the projection exercise, shocks to interest rates are generally of smaller magnitude. This feature may explain the somewhat larger coefficients on interest rate shocks in the tables.

suggest that this revision is linked into a new projection of potential output by shifting the previously projected level of potential output up or down in an essentially parallel fashion so that the shock has permanent effects.

Second, parallel revisions to the level of potential are consistent with smaller revisions to the output gap and potential growth, variables that play more prominent roles in communication. For communications purposes, it is preferable to focus on the main underlying signal of the state of the economy that indicates the extent of inflationary pressures. Large or frequent revisions to the recent history of the output gap or to projections of economic activity, particularly when reversed, would be undesirable. The historical mapping of a fraction of shocks into parallel shifts of potential output reduces the size of real-time revisions to the output gap and to projections of potential growth. In combination with communications about data revisions and uncertainty surrounding measures of potential output and the output gap, this may have provided a practical approach to dealing with real-time challenges of noisy and revised data.[21]

Finally, the pattern of revisions to projections of the output gap and *R*90 in response to GDP growth *shock1* may explain why there are only small effects of shocks on inflation. In particular, in general equilibrium, monetary policy responds (gradually according to the empirical results) to the revisions in the output gap projections. But with lags in the response of inflation to aggregate demand pressures, policy is "ahead of the curve" and attenuates inflationary implications. A similar outcome may occur with shocks to the exchange rate. In particular, projections of *R*90 at longer horizons respond positively to *EXCH* shocks (which are quite persistent, as evident in Table 1), possibly indicating slow pass-through of exchange rate movements to inflation, and therefore a delayed policy response to such shocks.

[21] Such a strategy is not unlike the strategy of using a measure of core inflation to indicate "underlying inflation" when a few components of the total CPI are subject to large transitory shocks.

CONCLUSION

The output gap plays a central role in monetary policy decisions and communications at the Bank of Canada. The methodology used to estimate and project potential output was designed to be consistent with the structure of the Bank's projection model (the QPM), allow estimates to be (flexibly) influenced by judgment and external structural estimates of trends, and incorporate information from a variety of sources to better disentangle supply and demand shocks. In practice, information sources that are external to the QPM, such as demographics or structural details of the Canadian labor market, are important drivers of the trend labor input component of potential output.

Analysis of revisions to real-time Bank of Canada staff economic projections reveals several interesting results. First, the similar size of typical revisions to projections of log potential output and the output gap suggest as much uncertainty about the trend as about the cycle. Second, real GDP shocks provided information about both the trend and the cycle. These shocks were parsed into permanent components that led to parallel shifts in projections of potential output and transitory components that led to persistent near-term revisions of the output gap that, with endogenous policy, dissipated over the projection horizon.

REFERENCES

Armstrong, John; Black, Richard; Laxton, Douglas and Rose, David. "The Bank of Canada's New Quarterly Projection Model, Part 2: A Robust Method for Simulating Forward-Looking Models." Bank of Canada Technical Report No. 73, Bank of Canada, February 1995; www.bankofcanada.ca/en/res/tr/1995/tr73.pdf.

Bank of Canada. *Monetary Policy Report: May 1995*. May 1995; www.bankofcanada.ca/en/mpr/pdf/mpr_apr_1995.pdf.

Bank of Canada. *Monetary Policy Report: May 1996*. May 1996; www.bankofcanada.ca/en/mpr/pdf/mpr_apr_1996.pdf.

Bank of Canada. *Monetary Policy Report: May 1999*. May 19, 1999; www.bankofcanada.ca/en/mpr/pdf/mpr_may_1999.pdf.

Bank of Canada. *Monetary Policy Report: November 2000*. November 9, 2000; www.bankofcanada.ca/en/mpr/pdf/mpr_nov_2000.pdf.

Barnett, Russell. "Trend Labour Supply in Canada: Implications of Demographic Shifts and the Increasing Labour Force Attachment of Women." *Bank of Canada Review*, Summer 2007, pp. 5-18; www.bankofcanada.ca/en/review/summer07/review_summer07.pdf.

Baxter, Marianne and King, Robert G. "Measuring Business Cycles: Approximate Bank-Pass Filters for Economic Time Series." *Review of Economics and Statistics*, November 1999, *81*(4), pp. 575-93.

Black, Richard; Laxton, Douglas; Rose, David and Tetlow, Robert. "The Bank of Canada's New Quarterly Projection Model, Part 1: The Steady-State Model: SSQPM." Technical Report No. 72, Bank of Canada, November 1994; www.bankofcanada.ca/en/res/tr/1994/tr72.pdf.

Butler, Leo. "The Bank of Canada's New Quarterly Projection Model, Part 4: A Semi-Structural Method to Estimate Potential Output: Combining Economic Theory with a Time-Series Filter." Technical Report No. 77, Bank of Canada, October 1996; www.bankofcanada.ca/en/res/tr/1996/tr77.pdf.

Coletti, Donald; Hunt, Benjamin; Rose, David and Tetlow, Robert. "The Bank of Canada's New Quarterly Projection Model, Part 3: The Dynamic Model: QPM." Technical Report No. 75, Bank of Canada, May 1996; www.bankofcanada.ca/en/res/tr/1996/tr75.pdf.

Côté, Denise and Hostland, Doug. "An Econometric Examination of the Trend Unemployment Rate in Canada." Working Paper No. 96-7, Bank of Canada, May 1996; www.bankofcanada.ca/en/res/wp/1996/wp96-7.pdf.

Crawford, Allan. "Trends in Productivity Growth in Canada." *Bank of Canada Review*, Spring 2002, pp. 19-32.

Dodge, David. Opening statement before the Standing Senate Committee on Banking, Trade and Commerce, November 29, 2001; www.bankofcanada.ca/en/speeches/2001/state01-4.html.

Guay, Alain and St-Amant, Pierre. "Do Mechanical Filters Provide a Good Approximation of Business Cycles?" Technical Report No. 78, Bank of Canada, November 1996; www.bankofcanada.ca/en/res/tr/1996/tr78.pdf.

Harvey, Andrew C. *Forecasting, Structural Time Series Models and the Kalman Filter*. New York: Cambridge University Press, 1989.

Hodrick, Robert J. and Prescott, Edward C. "Post-War U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit, and Banking*, February 1997, *29*(1), pp. 1-16.

Ip, Irene. "Labour Force Participation in Canada: Trends and Shifts." *Bank of Canada Review*, Summer 1998, pp. 29-52; www.bankofcanada.ca/en/review/1998/r983b.pdf.

King, Robert G. and Rebelo, Sergio, T. "Low Frequency Filtering and Real Business Cycles." *Journal of Economic Dynamics and Control*, 1993, *17*(1-2), pp. 207-31.

Kozicki, Sharon. "Multivariate Detrending under Common Trend Restrictions: Implications for Business Cycle Research." *Journal of Economic Dynamics and Control*, June 1999, *23*(7), pp. 997-1028.

Laxton, Douglas; Rose, David E. and Tetlow, Robert. "Is the Canadian Phillips Curve Non-Linear?" Working Paper 93-7, Bank of Canada, July 1993; www.douglaslaxton.org/sitebuildercontent/sitebuilderfiles/LRT3.pdf.

Laxton, Douglas and Tetlow, Robert. "A Simple Multivariate Filter for the Measurement of Potential Output." Technical Report No. 59, Bank of Canada, June 1992; www.douglaslaxton.org/ sitebuildercontent/sitebuilderfiles/LT.pdf.

Macklem, Tiff. "Information and Analysis for Monetary Policy: Coming to a Decision." *Bank of Canada*

Review, Summer 2002, pp. 11-18; www.bankofcanada.ca/en/review/2002/macklem_e.pdf.

Mise, Emi; Kim, Tae-Hwan and Newbold, P. "On Suboptimality of the Hodrick-Prescott Filter at Time-Series Endpoints." *Journal of Macroeconomics*, March 2005, *27*(1), pp. 53-67.

Orphanides, Athanasios and van Norden, Simon. "The Unreliability of Output Gap Estimates in Real Time." *Review of Economics and Statistics*, 2002, *84*(4), pp. 569-83.

Poloz, Stephen S. "The Causes of Unemployment in Canada: A Review of the Evidence." Working Paper 94-11, Bank of Canada, November 1994; www.bankofcanada.ca/en/res/wp/1994/wp94-11.pdf.

Poloz, Stephen; Rose, David and Tetlow, Robert. "The Bank of Canada's New Quarterly Projection Model (QPM): An Introduction." *Bank of Canada Review*, Autumn 1994, pp. 23-38; www.bankofcanada.ca/en/review/1994/r944a.pdf.

Rennison, Andrew. "Comparing Alternative Output Gap Estimators: A Monte Carlo Approach." Working Paper 2003-8, Bank of Canada, March 2003; www.bankofcanada.ca/en/res/wp/2003/wp03-8.pdf.

Stock, James H. and Watson, Mark W. "Testing for Common Trends." *Journal of the American Statistical Association*, December 1988, *83*(404), pp. 1097-107.

St-Amant, Pierre and van Norden, Simon. "Measurement of the Output Gap: A Discussion of Recent Research at the Bank of Canada." Technical Report No. 79, Bank of Canada, August 1997; www.bankofcanada.ca/en/res/tr/1997/tr79.pdf.

Thiessen, Gordon. "Monetary Policy and the Prospects for a Stronger Canadian Economy." Notes for remarks to the Canadian Association for Business Economics and the Ottawa Economics Association, Ottawa, Ontario, Canada. March 21, 1997; www.bankofcanada.ca/en/speeches/1997/sp97-3.html.

Commentary

Gregor W. Smith

parse: *v. tr.* resolve (a sentence) into its component parts and describe grammatically

In their thought-provoking and informative paper, Barnett, Kozicki, and Petrinec (2009) describe how the Bank of Canada used its quarterly projection model (QPM) between 1994 and 2005 to resolve changes in macroeconomic variables into their component parts. They make four distinct contributions by

(i) giving a history of the QPM,

(ii) describing how potential output was modeled with a multivariate filter that was outside the QPM and is still in use,

(iii) outlining the Bank of Canada's forecasting methods for potential output, and

(iv) illustrating the properties of multivariate forecast (or projection) errors and revisions.

In commenting on these contributions, I begin by looking at the history and forecasts of potential output as modeled by the Bank of Canada and then I draw attention to their findings concerning forecast revisions and forecast errors.

HISTORY AND FORECASTS OF POTENTIAL OUTPUT

During the 1990s the Bank of Canada began to model potential output with its extended multivariate filter (EMVF), a development that was ahead of its time. The Bank still uses this filter today. The filter is multivariate in the sense that it takes a range of indicators (e.g., the participation rate and the unemployment rate) as inputs in addition to output itself. The filter is extended in the sense that it uses economic information to define the output gap. This information includes restrictions requiring a common trend for some series or a positive correlation between the output gap and the inflation rate. Finally, the EVMF also is two sided, using both previous values of its input variables and subsequent values (or their forecasts, when potential output is being estimated for recent quarters).

The Bank's projection method thus involved two sets of parameters—one in the EVMF and another in the QPM. It is relatively easy to think of situations in which identifying parameters in the second component might depend on the parameterization in the first component. For example, the EVMF used parameter values that built in some smoothness in the series for potential output. The paper by Basu and Fernald (2009), in this issue, skeptically discusses the use of smoothness restrictions in defining potential output.

An alternative to the Bank's procedure would have been to smooth later in the process in the QPM. For example, using an unsmooth potential output series as an input in the QPM presumably would have led to a calibration of the QPM that involved smaller reactions to potential output or that used reactions to both current and lagged values of potential output, so that the smoothing

Gregor W. Smith is the Douglas D. Purvis Professor of Economics at Queen's University.

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 267-70.
© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

effectively occurred at this second stage. This particular sequence of calibrations or parameterizations is now history, for the Bank of Canada has replaced the QPM with the Terms-of-Trade Economic Model (ToTEM, as described by Fenton and Murchison, 2006), but the general point about possible interdependency remains.

The EVMF is a two-sided filter, which naturally raises the questions of how to replace some future values by forecasts, how to deal with revisions in data, and what to do near the end of a time-series sample. Here my own vote favors a one-sided approach using the Kalman filter to forecast, filter, and then smooth as data vintages accumulate. Of course, the two-sided filter can be written in terms of forecasts so that it is one sided. My suggestion is simply that such a process might be a clearer place to begin, because I do not interpret Barnett, Kozicki, and Petrinec as arguing for any special interest in the parameters of the two-sided version. Anderson and Gascon (2009), also in this issue, provide a comprehensive application of a one-sided approach to U.S. data.

Historical and forecast series for potential output at the Bank of Canada are based on *different* input series and restrictions. For example, forecasts of potential output use forecasts for the capital stock and total factor productivity, while historical estimates do not use these series. These two measures obviously serve different purposes. The Bank of Canada uses potential output and the output gap to convey the idea that accumulated events or output relative to the path of potential matter to current events such as the inflation rate. That communication can counteract the view that only the most recent growth rates of macroeconomic variables matter to the subsequent evolution of the economy. I would worry, though, that having different procedures for measuring historical, potential output, and forecasting current and future potential output might hinder the communication effort.

Barnett, Kozicki, and Petrinec also discuss the issue of the sensitivity of the Bank of Canada's measure of potential output to the endpoint. As they note, noisiness in the output gap limits its usefulness as a communication device. The alternative—the Kalman-filter approach—delivers a lower weight on the observation equation in preliminary data than in revised data to reflect this uncertainty. Therefore, that alternative approach again may be a natural framework for this issue. However, it is always possible that current measures of today's potential output and output gap are simply too noisy to be useful guides to policy, as Croushore (2009) suggests in his commentary in this issue.

FORECAST REVISIONS AND FORECAST ERRORS

Barnett, Kozicki, and Petrinec next focus on the properties of multivariate forecasts, also known as projections. Readers do not observe the exact model used to produce the forecasts because that model combines the QPM with estimates of trends. But studying forecasts is a natural way to evaluate the model anyway, so their study of forecast errors and revisions between 1993 and 2005 is welcome.

A key piece of notation is that x_t^v denotes a variable *for* quarter t and measured *at* quarter v (for vintage data). For given t, as v counts up, a switch occurs from forecasts to preliminary data and then to revised data. Unobserved variables involve only a succession of forecasts. Bearing in mind this sequence, I thus apply a gestalt switch to their Figures 1 and 2. I read them from right to left so that they describe the changes over time as the date t to which the forecasts apply first is approached then left behind.

To comment on their informative reporting, I use some notation. Let us define

$$\varepsilon_t^v = x_t^v - x_t^{v-1},$$

which is the one-vintage-apart change in the forecast. With $v \leq t$ this is a forecast revision; with $v > t$ it is a forecast error. This updating applies to three different types of variables: (i) those that are eventually observed and not subsequently revised (like the consumer price index [CPI]); (ii) those that are eventually observed but then revised (like gross domestic product [GDP]); and (iii) those that are never observed (like potential GDP). To help readers understand the forecast

process, Barnett, Kozicki, and Petrinec next provide three different types of statistics involving ε_t^v.

Standard Deviation Over Time

A first way to provide information about forecast errors and revisions is to document their variability over time using the standard deviation,

$$std_t\left(\varepsilon_t^v\right),$$

and then see how this varies with v. For most values of v, these standard deviations in potential GDP are roughly as large as those in revisions to actual GDP. But, of course, the revisions or errors in actual output and potential output are correlated, so the forecast revisions or errors for the output gap are much less volatile.

Correlation Across Horizons

A second way to study revisions or forecast errors is to look at their correlation over horizons:

$$corr_k\left(\varepsilon_{t+k}^v, \varepsilon_t^v\right).$$

This correlation naturally reflects the implied, underlying persistence. (Reporting correlations, if any, over time also would be interesting.)

Correlations of Forecast Errors Across Variables

A third, informative statistic is the correlation of revisions or forecast errors across macroeconomic variables. For example, using y to denote output and π to denote inflation, an interesting correlation is

$$corr_k\left(\varepsilon_{\pi k}^v, \varepsilon_{yt}^v\right).$$

This correlation is high between GDP growth and the output gap. It is low for inflation (or core inflation) and the output gap: 0.14 (or 0.05). The output gap could be measured or defined based on this correlation. I stress that that is *not* what the authors try to do. But since the Bank of Canada does use the output gap to try to communicate its views on inflation, it seems natural to test whether this "news" correlation is significantly different from zero.

Correlations of Forecast Errors Across Variables and Horizons

Perhaps the correlation at a longer horizon for inflation would be more interesting than the one between contemporaneous revisions. That would tell us how news about the output gap leads to immediate revisions in forecasts for subsequent inflation. The authors' Table 4 provides exactly this type of statistic:

$$corr\left(\varepsilon_{\pi t+k}^v, \varepsilon_{yt}^v\right).$$

Barnett, Kozicki, and Petrinec find a small, positive effect of GDP forecast errors on later inflation forecasts, an effect that is statistically significant at five to seven quarters (I am not sure of the units so cannot report on the economic significance). The policy interest rate rises too (as do market rates) but not enough to fully offset the effect of the change in the output gap on later, forecasted inflation.

Can we parameterize potential output so the output gap causes inflation? (Or can we test the causal role for a gap measured with a production function?) I think the answer is no, we cannot. At longer horizons, *nothing* should lead to revisions to inflation forecasts. Imagine a least squares regression like this:

$$\pi_{t+k}^t - 2.0 = \beta_0 + \beta_1 z_t^t.$$

In this regression, we should find that the coefficients are indistinguishable from zero for any variable z_t^t and any horizon, say, $k > 4$ quarters. Thus, this regression could not identify parameters of the output gap or potential output.

After 1995 the official inflation target (and average inflation rate) was 2 percent. Forecasts should equal that value at and beyond the horizon over which the policy interest rate has effect. Kuttner and Posen (1999), Rowe and Yetman (2002), and Otto and Voss (2009) have outlined this unforecastability of inflation departures from target under successful inflation targeting. So deviations from 2 percent in the Bank's inflation forecasts could reflect overflexible inflation targeting or an insufficient response of the policy interest rate. A role for the historical (two-sided) output gap in this regression would show that

alternative measurement to be misleading. But the response of the overnight interest rate (policy) perhaps could identify learning by the central bank about the output gap.

CONCLUSION

This commentary has followed Barnett, Kozicki, and Petrinec's article by beginning with how potential is and was measured in Canada and then turning to the properties of revisions and forecast errors. But perhaps this sequence could come full circle in the Bank of Canada's research: Studying the properties of forecast (projection) errors might well lead to changes in how the Bank measures potential output.

Under inflation targeting there is no information in inflation forecasts with which to test or identify lagged effects of potential output or the output gap on inflation. So, statistically, the output gap might be better thought of not as the thing that predicts inflation but rather as the thing to which the policy interest rate reacts and, implicitly, about which the Bank of Canada learns.

I conclude with a brief observation I would like to emphasize. Full credit goes to the Bank of Canada and its researchers for publicizing these data from past projections and documenting their properties. As this article shows, these data provide a rich source of insights into the tools used in monetary policy.

REFERENCES

Anderson, Richard G. and Gascon, Charles S. "Estimating U.S. Output Growth with Vintage Data in a State-Space Framework." Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 271-90.

Barnett, Russell; Kozicki, Sharon and Petrinec, Christopher. "Parsing Shocks: Real-Time Revisions to Gap and Growth Projections for Canada." Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 247-65.

Basu, Susantu and Fernald, John G. "What Do We Know (And Not Know) about Potential Output?" Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 187-213.

Croushore, Dean. Commentary on "Estimating U.S. Output Growth with Vintage Data in a State-Space Framework." Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 371-81.

Fenton, Paul and Murchison, Stephen. "ToTEM: The Bank of Canada's New Projection and Policy-Analysis Model." *Bank of Canada Review*, Autumn 2006, pp. 5-18.

Kuttner, Kenneth N. and Posen, Adam S. "Does Talk Matter After All? Inflation Targeting and Central Bank Behavior." Staff Report No. 88, Federal Reserve Bank of New York, October 1999; www.newyorkfed.org/research/staff_reports/sr88.pdf.

Otto, Glenn and Voss, Graham. "Tests of Inflation Forecast Targeting Models." Unpublished manuscript, Department of Economics, University of Victoria, 2009.

Rowe, Nicholas and Yetman, James. "Identifying a Policymaker's Target: An Application to the Bank of Canada." *Canadian Journal of Economics*, May 2002, *35*(2), pp. 239-56.

The Challenges of Estimating Potential Output in Real Time

Robert W. Arnold

Potential output is an estimate of the level of gross domestic product attainable when the economy is operating at a high rate of resource use. A summary measure of the economy's productive capacity, potential output plays an important role in the Congressional Budget Office (CBO)'s economic forecast and projection. The author briefly describes the method the CBO uses to estimate and project potential output, outlines some of the advantages and disadvantages of that approach, and describes some of the challenges associated with estimating and projecting potential output. Chief among these is the difficulty of estimating the underlying trends in economic data series that are volatile, subject to structural change, and frequently revised. Those challenges are illustrated using examples based on recent experience with labor force growth, the Phillips curve, and labor productivity growth. (JEL E17, E32, E62)

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 271-90.

Assessing current economic conditions, gauging inflationary pressures, and projecting long-term economic growth are central aspects of the Congressional Budget Office (CBO)'s economic forecasts and baseline projections. Those tasks require a summary measure of the economy's productive capacity. That measure, known as potential output, is an estimate of "full-employment" gross domestic product (GDP)—the level of GDP attainable when the economy is operating at a high rate of resource use. Although it is a measure of the productive capacity of the economy, potential output is not a technical ceiling on output that cannot be exceeded. Rather, it is a measure of *sustainable* output, where the intensity of resource use is neither adding to nor subtracting from short-run inflationary pressure. If actual output exceeds its potential level, then constraints on capacity begin to bind, restraining further growth and contributing to inflationary pressure. If output falls below potential, then resources are lying idle and inflation tends to fall.

In addition to being a measure of aggregate supply in the economy, potential output is also an estimate of trend GDP. The long-term trend in real GDP is generally upward as more resources—primarily labor and capital—become available and technological change allows more productive use of existing resources. Real GDP also displays short-term variation around that long-run trend, influenced primarily by the business cycle but also by random shocks whose sources are difficult to pinpoint. Analysts often want to estimate the underlying trend, or general momentum, in GDP by removing short-term variation from it. A distinct, but related, objective is to remove the fluctuations that arise solely from the effects of the business cycle.

Potential output plays a role in several areas associated with the CBO's economic forecast. In particular, we use potential output to set the level of real GDP in its medium-term (or 10-year)

Robert W. Arnold is principal analyst in the Congressional Budget Office.

© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

projections. In doing so, we assume that the gap between GDP and potential GDP will equal zero on average in the medium term. Therefore, the CBO projects that any gap that remains at the end of the short-term (or two-year) forecast will close during the following eight years. We also use the level of potential output as one gauge of inflationary pressures in the near term. For example, an increase in inflation that occurs when real GDP is below potential (and monetary growth is moderate) can probably be attributed to temporary factors and is unlikely to persist. Finally, potential output is an important input in computing the standardized-budget surplus or deficit, which the CBO uses to evaluate the stance of fiscal policy and reports regularly as part of its mandate.

THE CBO METHOD FOR ESTIMATING POTENTIAL OUTPUT

The CBO model for estimating potential output is based on the framework of a neoclassical, or Solow, growth model. The model includes a Cobb-Douglas production function for the nonfarm business (NFB) sector with two factor inputs, labor (measured as hours worked) and capital (measured as an index of capital services provided by the physical capital stock), and total factor productivity (TFP), which is calculated as a residual. NFB is by far the largest sector in the economy, accounting for 76 percent of GDP in 2007, compared with less than 10 percent for each of the other sectors. For smaller sectors of the economy, including farms, federal government, state and local government, households, and nonprofit institutions, simpler equations are used to model output. Those equations generally relate the growth of output in a sector to the growth of the factor input—either capital or labor—that is more important for production in that sector.[1]

To compute historical values for potential output, we cyclically adjust the factor inputs and then combine them using the production function. Cyclical adjustment removes the variation in a series that is attributable solely to business cycle fluctuations. Ideally, the resulting series will reflect not only the trend in the series, but also will be benchmarked to some measure of capacity in the economy and, therefore, can be interpreted as the potential level of the series.

For most variables in the model, we use a cyclic-adjustment equation that combines a relationship based on Okun's law with linear time trends to produce potential values for the factor inputs. Okun (1970) postulated an inverse relationship between the size of the output gap (the percentage difference between GDP and potential GDP) and the size of the unemployment gap (the difference between the unemployment rate and the "natural" rate of unemployment). According to that relationship, actual output exceeds its potential level when the unemployment rate is below the natural rate of unemployment and falls short of potential output when the unemployment rate is above its natural rate (Figure 1).

For the natural rate of unemployment, we use the CBO estimate of the non-accelerating inflation rate of unemployment (NAIRU). That rate corresponds to a particular notion of full employment—the rate of unemployment that is consistent with a stable rate of inflation. The historical estimate of the NAIRU derives from an estimated relationship known as a Phillips curve, which connects the change in inflation to the unemployment rate and other variables, including changes in productivity trends, oil price shocks, and wage and price controls. The historical relationship between the unemployment gap and the change in the rate of inflation appears to have weakened since the mid-1980s.[2] However, a negative correlation still exists; when the unemployment rate is below the NAIRU, inflation tends to rise, and when it exceeds the NAIRU, inflation tends to fall. Consequently, the NAIRU, while it is less useful for inflation forecasts, is still useful as a benchmark for potential output.

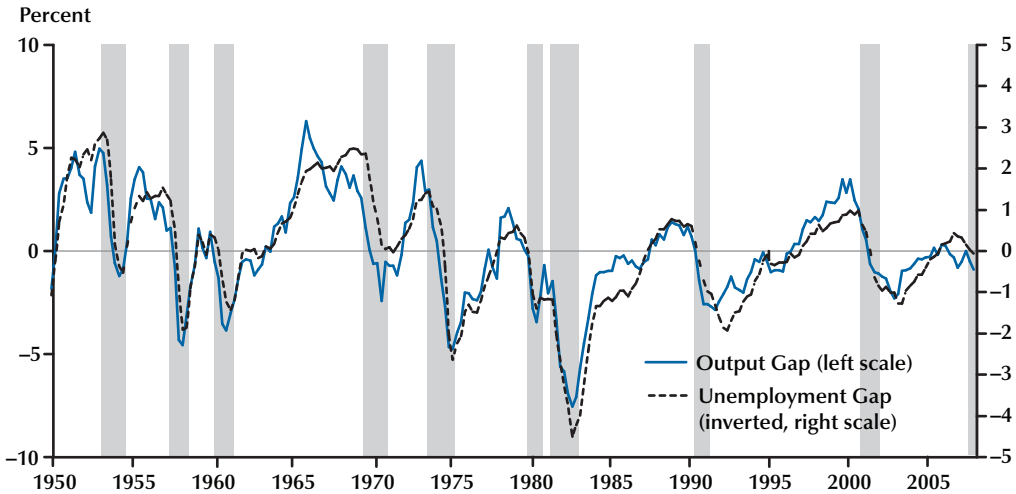
The assumption of linear time trends in the cyclic-adjustment equation implies that the poten-

[1] This section gives an overview of the CBO method. For a more complete description, see CBO (2001).

[2] For a description of the procedure used to estimate the NAIRU, see Arnold (2008).

Figure 1

Okun's Law: The Output Gap and the Unemployment Gap



NOTE: Gray bars in Figures 1, 2, 3, 5, 6, and 8 indicate recession as determined by the National Bureau of Economic Research.

tial version of each variable grows at a constant rate during each historical business cycle. Rather than constraining the potential series to follow a single time trend throughout the entire sample, the model allows for several time trends, each beginning at the peak of a business cycle. Defining the intervals of the time trends using full business cycles helps to ensure that the trends are estimated consistently throughout the historical sample. Most economic variables have distinct cyclical patterns—they behave differently at different points in the business cycle. Specifying breakpoints for the trends that occur at different stages of different business cycles (say, from trough to peak) would likely provide a misleading view of the underlying trend.

The cyclic-adjustment equation has the following form:

$$\log(X) = \text{Constant} + \alpha\left(U - U^*\right) + \beta_1 T_{1953} + \beta_2 T_{1957} + \ldots + \beta_8 T_{1990} + \varepsilon, \tag{1}$$

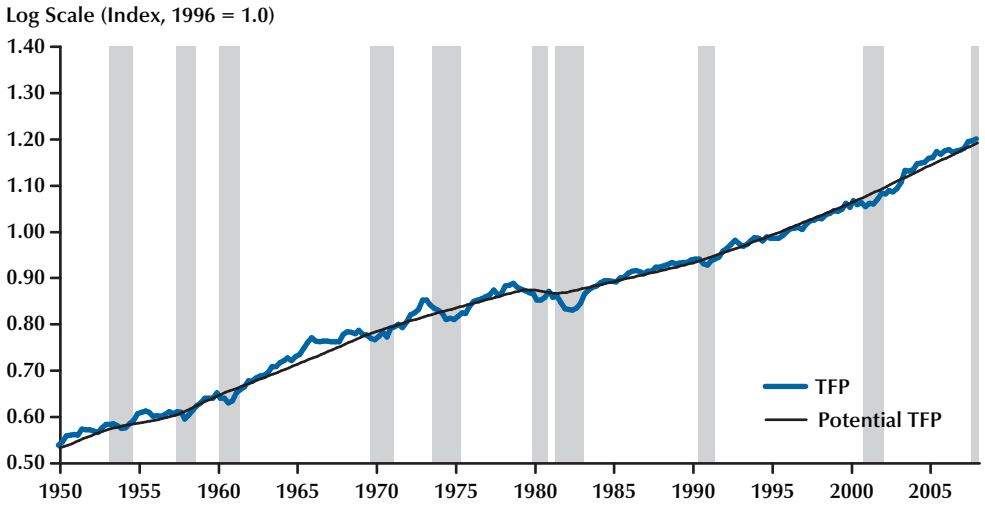
where X = the series to be cyclically adjusted,
U = unemployment rate,
U^* = NAIRU, and
T_i = zero until the business-cycle peak occurring in year i, after which it equals the number of quarters elapsed since that peak.

Equation (1), a piecewise linear regression, is estimated using quarterly data and ordinary least squares. Potential values for the series being adjusted are calculated as the fitted values from the regression, with U constrained to equal U^*. Setting the unemployment rate to equal the NAIRU removes the estimated effects of fluctuations in the business cycle; the resulting estimate gives the equation's prediction of what the dependent variable (X) would be if the unemployment rate never deviated from the NAIRU. An example of the results of using the cyclic-adjustment equation is illustrated in Figure 2, which shows TFP and potential TFP.

One question that arises is when to add a new trend break to the equation. Typically, we do not add a new breakpoint immediately after a business cycle peak because doing so would create, at least initially, a very short trend segment for the period after the peak. Such a segment would be subject to large swings as new data points were added

Figure 2

TFP and Potential TFP



to the sample because it was so short, at least initially. Because the final segment of the trend is carried forward into the projection, those swings would create instability to our medium-term projections. Consequently, we typically wait until a full business cycle has concluded before adding a new break to the trend. For example, the model does not yet include a break in 2001, though the addition of one appears to be increasingly likely.

Equation (1) is used for most, but not all, inputs in the model. One important exception is the capital input, which does not need to be cyclically adjusted to create a "potential" level because the unadjusted capital input already represents its potential contribution to output. Although use of the capital stock varies greatly during the business cycle, the potential flow of capital services is always related to the total size of the capital stock, not to the amount currently being used. Other exceptions include several variables of lesser importance that do not vary with the business cycle. Those series are smoothed using the Hodrick-Prescott filter.

As noted earlier, the method for computing historical values of potential output in the other sectors of the economy differs slightly from that used for the NFB sector. In general, the approach is to express real GDP in each of the other sectors as a function of the primary factor input (either labor or capital) in that sector and the productivity of that input. The potential levels of the primary input and its productivity are cyclically adjusted using an analog to equation (1) and then combined to estimate potential output in that sector. The list below describes how each sector is modeled.

- Farm sector: Potential GDP in this sector is modeled as a function of potential farm employment and potential output per employee.
- Government sector: Potential GDP in this sector is the sum of potential GDP in the federal government and state and local governments. Potential GDP at each level of government equals the sum of the compensation of general government employees (adjusted to potential) and government depreciation. Compensation is modeled as a function of total employment, and compensation per employee and depreciation is modeled as a function of the government capital stock.
- Nonprofit sector: Potential GDP in this sector is modeled as a function of potential

nonprofit employment and potential output per employee.

- Household sector: Although some of the GDP in the household sector consists of the compensation of domestic workers, the majority is composed of imputed rent on owner-occupied housing. As such, output in this sector is composed of a stream of housing services provided almost entirely from the capital stock. Potential GDP in this sector is modeled as a function of the stock of owner-occupied housing and an estimate of the productivity of that stock. Similar to the capital input in the NFB sector, the housing capital stock is not adjusted to potential because the unadjusted stock reflects the potential contribution to output.

For projections of potential output, the same framework is used for these sectors as is used for the NFB sector. Given projections of several exogenous variables—of which potential labor force, potential TFP growth, and the national saving rate are the most important—the growth model computes the capital stock endogenously and combines the factor inputs into an estimate of potential output. In most cases, projecting the exogenous variables is straightforward: The CBO generally extrapolates the trend growth rate from recent history through the 10-year projection period. However, the projections for some exogenous variables, most notably the saving rate, are taken from the CBO economic forecast.

Advantages and Disadvantages of the CBO Method

The CBO method for estimating and projecting potential output has several key advantages. First, it looks explicitly at the supply side of the economy. Potential output is a measure of productive capacity, so any estimate of it is likely to benefit from explicit dependence on factors of production. For example, if growth in the available pool of labor increases, then this method will show an acceleration in potential output (all other things being equal). With our approach, an increase in investment spending would also be reflected in faster growth in productive capacity.

Another advantage of a growth model is that it allows for a transparent accounting of the sources of growth. Such a growth-accounting exercise, which divides the growth of potential GDP into the contributions from each of the factor inputs, is especially useful when explaining the factors that caused a change to CBO projections. A growth-accounting exercise for our current projection is shown in Table 1.[3] The table displays the growth rates of potential output and its components for the overall economy and the NFB sector. Note that the growth rates of the factor inputs (top and middle panels of the table) are not weighted; they do not sum to the growth in potential output.

A third advantage of using a growth model to calculate potential output is that it supplies a projection for potential output that is consistent with the CBO projection for the federal budget. That consistency allows the CBO to incorporate the effects of changes in fiscal policy into its medium-term (10-year) economic and budget projections. Fiscal policy has obvious effects on aggregate demand in the short run, effects that are reflected in our short-term forecast. However, fiscal policy will also influence the growth in potential output over the medium term through its effect on national saving and capital accumulation. Because the growth model explicitly includes capital as a factor of production, it captures that effect.

Table 1 also shows the contribution of each factor input to the growth of potential output in the NFB sector by weighting each input's growth rate by its coefficient in the production function. The sum of the contributions equals the growth of potential output in the NFB sector. Computing the contributions to growth highlights the sources of any quickening or slowdown in growth. For example, the CBO estimates that potential output in the NFB sector grew at an average annual rate of 3.3 percent during the 1982-90 period and 3.5 percent during the 1991-2001 period. That acceleration can be attributed to faster growth in the capital input (which contributed 1.2 percentage points to the growth of potential output during the first period and 1.4 percentage points in the

[3] See CBO (2008).

Table 1

Key Assumptions in the CBO's Projection of Potential Output

	Average annual growth (%)						Projected average annual growth (%)		
	1950-73	1974-81	1982-90	1991-2001	2002-2007*	Total, 1950-2007*	2008-2013	2014-2018	Total, 2008-2018
Overall economy									
Potential output	3.9	3.2	3.1	3.1	2.7	3.4	2.5	2.4	2.4
Potential labor force	1.6	2.5	1.6	1.2	1.1	1.6	0.8	0.5	0.7
Potential labor force productivity†	2.3	0.7	1.4	1.9	1.6	1.8	1.6	1.9	1.7
Nonfarm business sector									
Potential output	4.0	3.6	3.3	3.5	3.0	3.6	2.8	2.8	2.8
Potential hours worked	1.4	2.3	1.7	1.1	1.0	1.5	0.7	0.4	0.6
Capital input	3.8	4.2	4.1	4.6	2.5	3.9	2.9	3.5	3.2
Potential TFP	1.9	0.7	0.9	1.3	1.5	1.4	1.4	1.4	1.4
Potential TFP excluding adjustments	1.9	0.7	0.9	1.3	1.3	1.4	1.3	1.3	1.3
TFP adjustments	0.0	0.0	0.0	0.1	0.2	‡	0.1	0.1	0.1
Price measurement§	0.0	0.0	0.0	0.1	0.1	‡	0.1	0.1	0.1
Temporary adjustment¶	0.0	0.0	0.0	‡	‡	‡	0.0	0.0	0.0
Contributions to the growth of potential									
Output (percentage points)									
Potential hours worked	1.0	1.6	1.2	0.8	0.7	1.0	0.5	0.3	0.4
Capital input	1.1	1.3	1.2	1.4	0.8	1.2	0.9	1.0	1.0
Potential TFP	1.9	0.7	0.9	1.3	1.5	1.4	1.4	1.4	1.4
Total contributions	4.0	3.6	3.3	3.5	2.9	3.6	2.8	2.8	2.8
Potential labor productivity in the NFB sector#	2.6	1.3	1.6	2.4	1.9	2.1	2.1	2.3	2.2

NOTE: Data are for calendar years. Numbers in the table may not add up to totals because of rounding.
*Values as of August 22, 2008.
†The ratio of potential output to the potential labor force.
‡Between zero and 0.05 percent.
§An adjustment for a conceptual change in the official measure of the GDP chained price index.
¶An adjustment for the unusually rapid growth of TFP between 2001 and 2003.
#The estimated trend in the ratio of output to hours worked in the NFB sector.

SOURCE: CBO.

second) and faster growth of potential TFP (which contributed 0.9 and 1.3 percentage points to the growth of potential output during the two periods, respectively). Faster growth in those two factors more than offset a slowdown in potential hours worked between the two periods. This point is addressed later.

Fourth, by using a disaggregated approach, the CBO method can reveal more insights about the economy than a more-aggregated model would. For example, the model calculates the capital input to the production function as a weighted average of the services provided by seven types of capital. Those data indicate a shift over the past few decades to capital goods with shorter service lives: A larger share of total fixed investment is going to producers' durable equipment (PDE) relative to structures, and a larger share of PDE is going to computers and other information technology (IT) capital. Because shorter-lived capital goods depreciate more rapidly, the shift toward PDE and IT capital increases the share of investment dollars used to replace worn-out capital and tends to lower net investment and the capital input. Shorter-lived capital goods are also more productive per year of service life than those that last longer and are therefore weighted more heavily in the growth model's capital input. A model that ignores the capital input or that does not disaggregate capital is likely to miss both of those effects.

On the negative side, the simplicity of our model could be perceived as a drawback. The model uses some parameters—most notably, the coefficients on labor and capital in the production function—that are imposed rather than econometrically estimated. Although that approach is standard practice in the growth-accounting literature (in part because it has empirical support), it is tantamount to assuming the magnitude of the contribution that each factor input makes to growth. With such an approach, the magnitude of that contribution will not change from year to year as the economy evolves, as it would in an econometrically estimated model. Moreover, it requires some strong assumptions that may not be consistent with the data.

A second disadvantage of using a growth model to estimate potential output is that including the capital stock introduces measurement error. Most economic variables are subject to measurement error, but the problem is particularly acute for capital, for two basic reasons. First, measuring the stock of any particular type of capital is difficult because depreciation is hard to define or measure. Purchases of plant and equipment can be tallied to produce a historical series for investment, but no corresponding source of data exists for depreciation. Second, even if accurate estimates of individual stocks were available, aggregating them into a single index would be difficult because capital is heterogeneous, differing with respect to characteristics such as durability and productivity.[4]

A third point of contention regarding the CBO approach is the use of deterministic time trends to cyclically adjust many variables in the model. Some analysts assert that relying on fixed time trends provides a misleading view of the cyclical behavior of some economic time series. They argue, on the basis of empirical studies of the business cycle, that using variable rather than fixed time trends is more appropriate for most data series.[5] However, the evidence on this point is mixed—it is very difficult to determine whether the trend in a data series is deterministic or stochastic using existing econometric techniques—and the methods used to estimate stochastic trends often yield results that are not useful for estimating potential output. That is, stochastic trends tend to produce estimates of the output gap that are not consistent with other indicators of the business cycle.

Fourth, the CBO growth model is based on an estimate of the amount of slack in the labor market, which in turn requires an estimate of the natural rate of unemployment or the NAIRU. Such

[4] The CBO capital input uses capital stock estimates (and the associated assumptions about depreciation) from the Bureau of Economic Analysis and uses an aggregation equation that is based on the approach used by the Bureau of Labor Statistics (BLS) to construct the capital input that underlies the multifactor productivity series. The CBO estimate of the capital input is quite similar to that calculated by the BLS.

[5] See, for example, Stock and Watson (1988).

estimates are highly uncertain. Few economists would claim that they can confidently identify the current NAIRU to within a percentage point. Our method is not very sensitive to possible errors in the average level of the estimated NAIRU, but it is sensitive to errors in identifying how that level changes from year to year.

Finally, the CBO model does not contain explicit channels of influence for all major effects of government policy on potential output. For example, it does not include an explicit link between tax rates and labor supply, productivity, or the personal saving rate; nor does it include any link between changes in regulatory policy and those variables. However, that does not mean that the model precludes a relationship between policy changes and any of those variables. If a given policy change is estimated to be large enough to affect the incentives governing work effort, productivity, or saving, then those effects can be included in our projection or in a policy simulation by adjusting the relevant variable in the model. For example, changes in marginal tax rates have the potential to affect labor supply. Because the Solow model does not explicitly model the factors that affect the labor input, our model includes a separate adjustment to incorporate such effects. Indeed, for the past several years, such an adjustment has been included in our model to account for the effects on the labor supply of the scheduled expiration in 2011 of the tax laws passed in 2001 and 2003. The structure of our model makes it easier to isolate (and incorporate) the effects of such policy changes than would be the case with a time-series–based model.

CHALLENGES ASSOCIATED WITH ESTIMATING AND PROJECTING POTENTIAL OUTPUT

Potential output plays a key role in the CBO economic forecast and projection. Perhaps the two most important are estimating the output gap (percentage difference between GDP and potential GDP) and providing a target for the 10-year projection of GDP. Important challenges are associated with both roles.

Challenges Associated with Estimating the Output Gap

Any method used to estimate the trend in a series, including potential output, is subject to an "end-of-sample" problem, which means that estimating the trend is especially difficult near the end of a data sample. In the case of the output gap, this is usually the period of greatest interest. Three examples from the period since 2000 illustrate the difficulties associated with estimating the level of potential output at the end of the sample period.

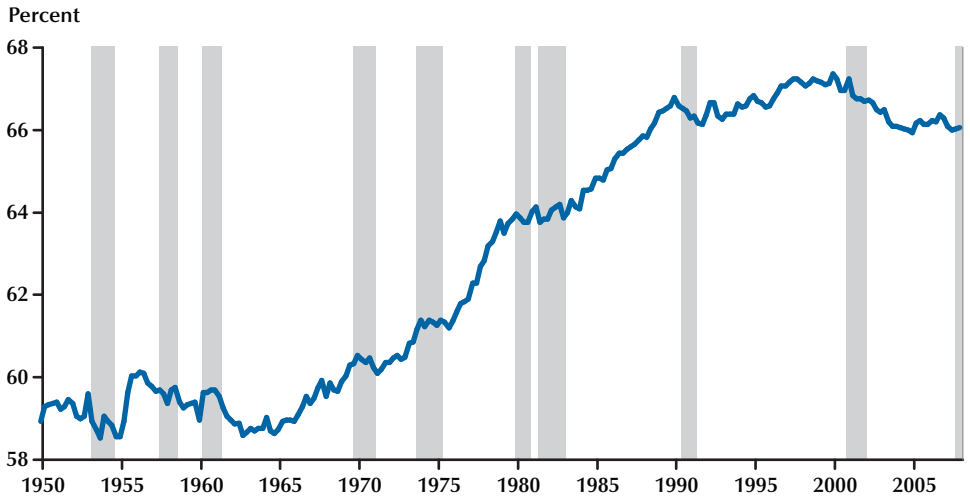
Potential Labor Force. Fundamentally, the amount of hours worked in the economy is determined by the size of the labor force, which, in turn, is largely influenced by two factors: growth in the population and the rate of labor force participation. Neither of those series is especially sensitive to business cycle fluctuations, but both are subject to considerable low-frequency variation. The discussion here focuses on how the rate of labor force participation has changed during recent years and how we have modified the CBO labor force projections as a result.

After a long-running rise that started in the early 1960s, the labor force participation rate plateaued at about 67 percent of the civilian population during the late 1990s, declined sharply between 2000 and 2003, and varied in a narrow range near 66 percent between 2003 and 2008 (Figure 3). Had that decline in the participation rate not occurred, the labor force would have had approximately 2.3 million more workers in 2008 than it actually did.

To assess its impact on potential output, the challenge during the early 2000s was to determine whether the decline in the participation rate was cyclical (i.e., workers had dropped out of the labor force because their prospects of getting a job were dim) or structural (i.e., prospective labor force participants had weighed the alternatives and found that options such as education, retirement, or child-rearing were more attractive). If the decline were due to cyclical reasons, then the dip in participation should not be reflected in the estimate of potential labor force. If the decline were due to structural reasons, however, then the

Figure 3

Labor Force Participation Rate



estimates of potential labor force and potential output should be lowered to reflect the decreased size of the potential workforce.

The drop in the participation rate also complicated the interpretation of movements in the unemployment rate, which peaked at 6.1 percent in mid-2003 and declined thereafter. During 2006 and 2007, the unemployment rate was below 5 percent, which suggested considerable tightness in the labor market. However, the decline in the participation rate implied that there existed a pool of untapped labor that could have been drawn into the workforce had there been a significant speedup in the pace of job creation. Consequently, at that time, the unemployment rate probably understated the degree of slack that existed in the labor market. Indeed, in the early stages of the expansion following the 2001 recession, we projected that the participation rate would recover as job creation picked up. It never did though, and the CBO has since concluded that the decline in the participation rate was more structural than cyclical.[6]

[6] That conclusion was based on an analysis of the factors affecting the participation rates of various demographic subgroups in the population; see CBO (2004).

Potential NFB Employment. The second challenge is associated with the behavior of employment since the end of the 2001 recession and its implications for the estimate of potential hours worked in the NFB sector. One striking feature of the economic landscape since the 2001 business cycle trough is very weak growth in employment, especially for measures derived from the Bureau of Labor Statistics' establishment survey. For example, since the trough in the fourth quarter of 2001, growth in nonfarm payroll employment averaged 0.8 percent at an annual rate, which means that payrolls were roughly 5 percent higher in the second quarter of 2008 than they were at the end of the 2001 recession. However, based on patterns in past cycles, one would have expected much faster growth in payroll employment—2.4 percent on average—and a much higher level of employment—17 percent higher than its trough value—by the second quarter of 2008 (Figure 4).

A similar pattern holds for employment in the NFB sector (which differs from the headline payroll number by excluding employees in private households and nonprofit institutions and including proprietors). In the second quarter of 2008, NFB employment was about 4 percent above

Figure 4

Payroll Employment in the Current Expansion Compared with an "Average" Cycle

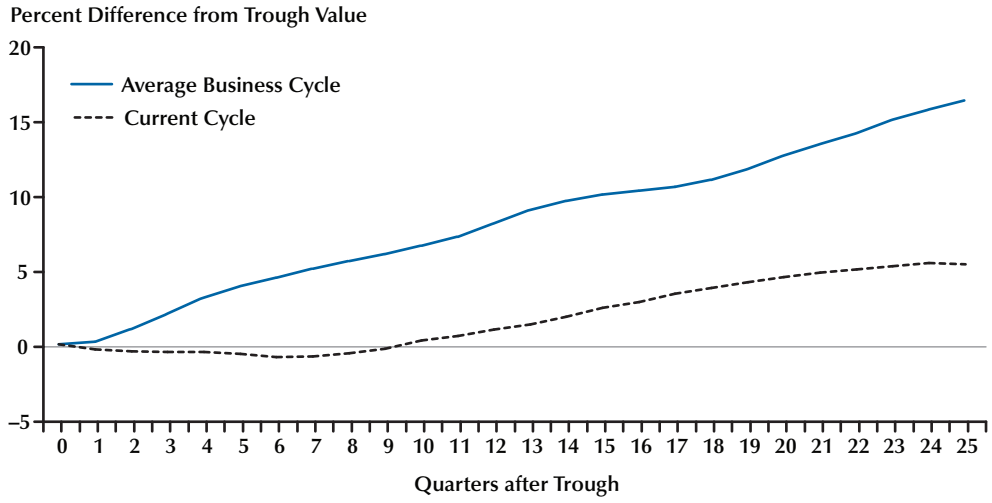


Figure 5

NFB Sector Employment as a Percent of the Civilian Labor Force

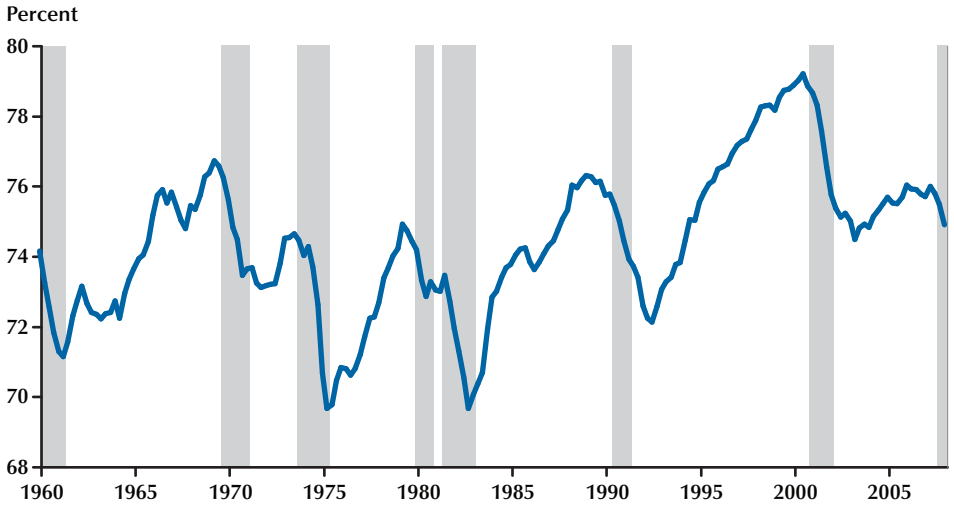
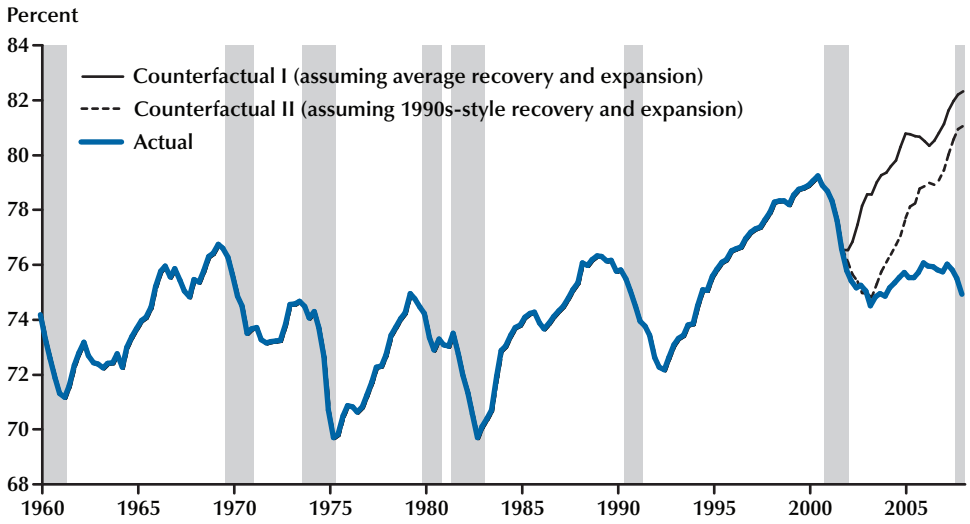


Figure 6

NFB Employment as a Percent of the Civilian Labor Force and Two Counterfactual Paths



its level at the end of the 2001 recession. Had it grown according to the pattern seen in a typical business cycle expansion, it would have been about 15 percent higher than its level at the trough of the recession.

The behavior of NFB employment since the business cycle peak in 2001 also looks very unusual when viewed from another perspective. When measured as a share of the labor force (which controls for the decline in the rate of labor force participation), NFB employment barely grew during the expansion that followed the 2001 recession (Figure 5). This is extremely unusual on two counts. First, it departs from the very strong procyclical pattern seen in most recovery and expansion periods. Typically, NFB sector employment grows much faster than the labor force during business cycle expansions, which causes a rapid increase in its share. Second, the recent behavior breaks with the long-standing upward trend in the NFB share of the labor force. Since roughly the mid-1970s, trend growth in NFB employment has exceeded trend growth in the labor force on average, leading to a steady increase in the share. Examining the peaks is a rough-and-ready way to control for business cycle variation: The share increased from about 74 percent in 1973, to just under 75 percent in 1979, to just over 76 percent in 1989, to just under 80 percent in 1999.

After the trough in 2001, the share of NFB employment declined for another two years and then increased somewhat but not anything like a normal cyclical rebound. The reasons for this behavior are not fully clear—shifts of employment to other sectors, including government and non-profit institutions, can explain only part of the shortfall—but it has important implications for the estimate of potential employment and hours worked. Specifically, the estimate of potential employment in the NFB sector is much lower than it would have been had actual NFB employment followed a more typical cyclical pattern since 2001.

To illustrate this point, consider what the NFB employment share would have looked like had it followed a more typical cyclical pattern. Figure 6 shows NFB employment as a share of the labor force along with two counterfactual paths for the share. The thin solid line shows the evolution of the NFB employment share had it grown since 2001 at the same rate as an "average" historical expansion. That path embodies much stronger

employment growth than does the actual path and would imply a much higher level of potential employment as well. Arguably, that path is too strong, given that employment growth has been sluggish in the recoveries that followed the past two recessions. So the figure includes a second counterfactual path (dotted line) showing the evolution of the NFB employment share had it grown as it did during the expansion of the 1990s. It too implies much stronger employment growth than actually occurred.

For the first few years of the current business cycle, it was reasonable to expect a typical rebound in the NFB employment share, even if it was delayed relative to past expansions. If so, then the period of sluggish growth in NFB employment could be interpreted as a cyclical pattern and would not necessarily imply that the level of potential NFB employment was lower. However, as the period of sluggish growth grew longer and in light of the possibility of a business cycle peak in early 2008, the position that NFB employment would eventually rebound became increasingly untenable.

Instead, it seems increasingly likely that NFB employment will merely match the growth in the labor force in the future, rather than grow at a faster pace. One implication of that interpretation is that the experience of the late 1990s, when the NFB employment share of the labor force was very high, was unusual and is unlikely to be repeated.

Changes in the Phillips Curve and NAIRU. As noted previously, the natural rate of unemployment is an important input in the CBO model. It serves as the benchmark used to estimate the potential values of the factor inputs and, consequently, potential output. Any uncertainties associated with the size of the unemployment gap, or difference between the unemployment rate and the natural rate, will translate directly into uncertainty about the size of the output gap.

Our estimate of the natural rate, known as the NAIRU, is based on a standard Phillips curve, which relates changes in inflation to the unemployment rate, expected inflation, and various supply shock variables. In particular, the NAIRU estimate relies on the existence of a negative correlation between inflation and unemployment: If inflation tends to rise when the unemployment rate is low and tends to fall when the unemployment rate is high, then there must be an unemployment rate at which there is no tendency for inflation to rise or fall. This does not mean that the rate is stable or that it is precisely estimated, just that it must exist.

However, during the past 20 or so years, significant changes in how the economy functions have affected the relationship between inflation and unemployment and, consequently, estimates of the Phillips curve and the NAIRU. Most notably, the rate of inflation has been lower and much less volatile since the mid-1980s, a phenomenon often referred to as the Great Moderation. At the same time, the unemployment rate has trended downward, which suggests that the natural rate of unemployment has declined also. Researchers had identified several factors that would be expected to lower the natural rate, including the changing demographic composition of the workforce, changes in disability policies, and improved efficiency of the labor market's matching process. Based on internal evaluation of those factors, the CBO began to lower its estimate of the NAIRU for the period since 1990, overriding the econometric estimate at that time.[7]

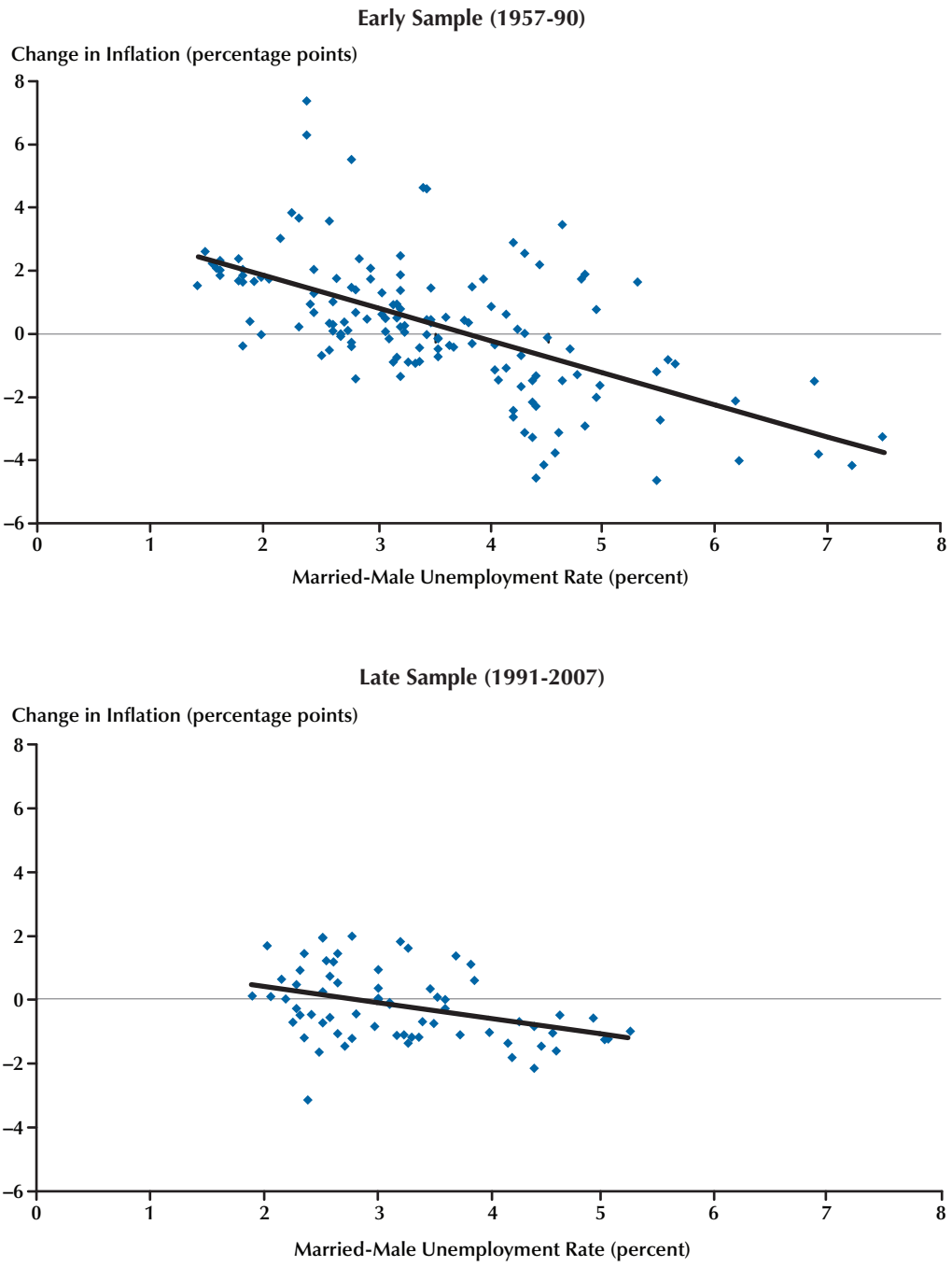
More recent Phillips curve estimates are consistent with the hypothesis that a change occurred sometime during the past 20 or so years. In a recent working paper, I presented regression results from estimates of several Phillips curve specifications that suggested the presence of significant structural change since the mid-1980s.[8] Using the full data sample, from 1955 through 2007, the equations' performance appeared to be satisfactory. They fit the data well and their estimated coefficients had the correct sign, were of reasonable magnitude, and were statistically significant. However, the full-sample regressions masked evidence of a breakdown in performance that began during the mid-1980s. Estimation results from equations that allowed for structural change indicated that the fit of the equations dete-

[7] That analysis was later summarized in a CBO paper; see Brauer (2007).

[8] See Arnold (2008).

Figure 7

Married-Male Unemployment and the Change in Inflation



NOTE: The change in inflation is defined as the difference between the quarterly rate of inflation in the personal consumption expenditure (PCE) price index and a 24-quarter moving average of PCE inflation.

riorated and that the coefficients were smaller and less statistically significant during the latter part of the data sample than they were during the earlier part. In general, the results suggest that the NAIRU is lower now than it had been during the period from 1955 through the mid-1980s, a conclusion consistent with evidence from the labor market suggesting a decline in the natural rate. The results also indicate that the Phillips curve has become less useful for predicting inflation.

However, the relationship between inflation and unemployment, though not as strong as it once was, has not collapsed completely. Consider Figure 7, which plots changes in a measure of unanticipated inflation against the married-male unemployment rate. The top panel shows data from the 1957-90 period, while the bottom panel shows data from 1991 through 2007.[9] Comparing the two panels reveals four features of the latter period. First, both graphs show a negative correlation between the two series, so there still appears to be a tradeoff between inflation and unemployment. Second, the point at which the regression line intersects the horizontal-axis intercept has moved to the left in the second panel, which is consistent with the idea that the NAIRU is lower now than it had been earlier. Third, the slope of the trend line is lower during the second part of the sample, which suggests that the inflation-unemployment tradeoff is somewhat flatter during the second period (i.e., inflation is less responsive to changes in the unemployment rate). Fourth, much less variation has occurred in both inflation and unemployment during the past 20 or so years than previously.

What do these observations imply for the estimate of potential output? The first observation—that a negative correlation still exists—means that the unemployment rate is still consistent with a stable rate of inflation. The second observation—that the NAIRU has declined—implies that the level of potential output is higher than it would have been had the NAIRU been constant. This observation also serves as a reminder that structural change in macro equations is a fact of life. It is important to monitor such equations continually to identify how economic events will affect their conclusions. The final two observations imply that Phillips curves, and by extension the NAIRU and potential output, are less useful indicators of inflationary pressure than they once were.

Challenges Associated with Projecting Potential Output

Potential output is used for more than gauging the state of the business cycle. It is also used to set the path for real GDP in the 10-year forecast that underlies the CBO budget projections. A separate set of challenges is associated with projecting the variables that underlie our estimate of potential output.

Projecting Labor Productivity I: The Late-1990s' Acceleration. Labor productivity growth during the late 1990s provides an important example of the challenges associated with projecting potential GDP.[10] The broad outlines of the story are familiar: After a long period of sluggish growth, labor productivity accelerated sharply during the second half of the 1990s and continued to grow rapidly during the 2000s. Moreover, the upswing was substantial. Trend growth in labor productivity averaged 2.7 percent between the end of 1995 and the middle of 2008, considerably faster than the 1.4 percent pace from 1974 to 1995 (Figure 8). Had it followed that pre-1996 trend of 1.4 percent instead of growing as it did, labor productivity would be 15 percent lower than it is today. Furthermore, if the 2.7 percent trend is sustained over the next decade, then the level of real GDP will be nearly 30 percent higher in 2018 than the level that would have resulted from the pre-1996 rate of growth.

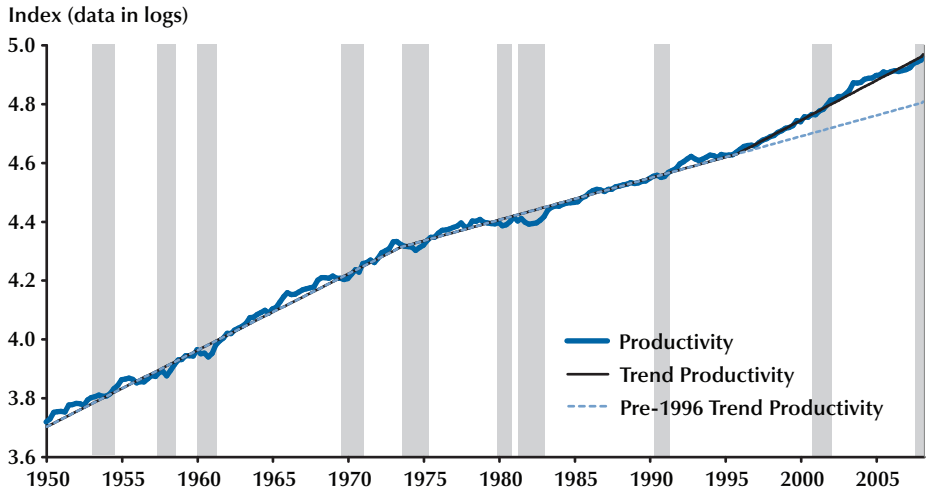
One problem for forecasters was that the productivity acceleration was largely unexpected. In the mid-1990s, few analysts anticipated such a dramatic increase in the trend rate of growth.

[9] The working paper estimated Phillips curve equations using different price indices and used Chow tests to determine when the structural break occurred in each equation. For the personal consumption expenditures price index, the break was found in 1991. The married-male unemployment rate was used in the estimation because it is better insulated from demographic shifts than the overall unemployment rate.

[10] In our model, we actually project potential TFP—the projection for potential labor productivity is implied by the projections for potential TFP and capital accumulation.

Figure 8

Labor Productivity Growth and Trend (1950-2008)



In January 1995, for example, the CBO projected that labor productivity growth would average 1.3 percent annually for the 1995-2000 period, a pace similar to the average for the prior 20 years. The Clinton administration and the Blue Chip Consensus of private forecasters projected similar rates of growth.

Another problem for forecasters was that the productivity surge in the late 1990s went unrecognized until very late in the decade for two basic reasons. First, labor productivity is fairly volatile, with growth rates that can swing widely from quarter to quarter. As a result, a period of two or three years is a short window within which to discern a new trend. Moreover, the acceleration followed a period of subpar growth (productivity growth averaged only 0.22 percent annually between the end of 1992 and 1995:Q3); so, initially, the faster growth appeared to be just making up lost ground rather than establishing a new, higher trend growth rate. The postwar data sample includes several episodes of faster- or slower-than-trend productivity growth that were later reversed.

Second, early vintages of productivity data for the late 1990s proved to be understated and, therefore, painted a misleading picture of the productivity trend. Only after several revisions did a stronger pattern emerge. Using real-time data culled from our forecast databases, Table 2 shows that data available in 1996, 1997, and 1998 showed only a small rise in productivity growth starting in late 1995. For example, data available in early 1997 showed labor productivity growing by only 0.3 percent on average from 1995:Q4 through 1996:Q3. The story changes markedly using currently available data: Labor productivity growth for that period was actually 3 percent.

A similar case holds for 1998 and 1999. Data available in January 1998 showed labor productivity growth averaging 1.8 percent between 1995:Q4 and 1997:Q3. That rate has since been revised upward by 0.6 percentage points, to 2.4 percent. The growth rate for the three-year period ending in 1998:Q3 also has been revised from 2.0 percent (using data from early 1999) to 2.5 percent (using currently available data).

The information in Table 2 highlights an important point. Productivity data are revised frequently, and the revisions can be large enough to alter analyses of trends in productivity growth. Indeed, after being revised upward several times during the late 1990s, productivity data have been revised downward somewhat during recent years.

Table 2

Changes in Estimates of Average Annual Growth Rate for Labor Productivity

Date of forecast	Period	Average annual rate of growth (%) Initial estimate (using original data)	Current estimate (using current data)	Revision (percentage points)
January 1997	1995:Q4–1996:Q3	0.3	3.0	2.7
January 1998	1995:Q4–1997:Q3	1.8	2.4	0.6
January 1999	1995:Q4–1998:Q3	2.0	2.5	0.5
January 2000	1995:Q4–1999:Q3	2.7	2.4	–0.3

NOTE: Each forecast is based on productivity data that extend through the third quarter of the previous year. Numbers in the table may not add up to totals because of rounding.

SOURCE: CBO based on data from the BLS.

In January 2000, labor productivity growth for 1995:Q4 to 1999:Q3 was estimated at 2.7 percent; that estimate has since been revised to 2.4 percent.

The revisions to productivity data highlight the difficulty in recognizing a change in the underlying trend growth rate and suggest that we should be circumspect about data series until they have undergone revision. This is especially true if the data show a shift in trend (as in the late 1990s) or if they are not consistent with other economic indicators.

Projecting Labor Productivity II: Shifting Sources of Growth. Another aspect of labor productivity growth during the past decade—a shift in its sources—has complicated the analysis of trends and made projections difficult. With our model we can easily divide the growth in labor productivity into two components: capital deepening (increases in the amount of capital available per worker) and TFP. Capital per worker can rise over time not only because investment provides more capital goods for workers to use, but also because the quality of those goods improves over time and investment can shift from assets with relatively low levels of productivity (e.g., factories) to those with higher productivity levels (e.g., computers). Because TFP is calculated as a residual—the growth contributions of labor and capital are subtracted from the growth in output—any growth in labor productivity that is not attributed to capital deepening will be assigned to TFP.

With this in mind, the contributions of capital deepening and TFP to the growth in labor productivity since 1995 can be calculated. The results of such a growth-accounting exercise are shown in Table 3. Those results show that capital deepening was the primary source of the surge in labor productivity growth in the late 1990s and that faster TFP growth was the primary source of productivity growth during the period after the business cycle peak in 2001. Between the early (1991-95) and the late (1996-2001) part of the past decade, labor productivity growth stepped up from about 1.5 percent, on average, to 2.5 percent per year. Growth in capital per worker accounted for 80 percent (0.8 percentage points) of that 1-percentage-point increase, according to our estimates. Faster TFP growth was responsible for the rest of the step-up in productivity growth, or about 0.2 percentage points.

Since the 2001 recession, however, the sources of labor productivity growth have completely reversed. Business investment fell substantially in 2001 and 2002 and remained weak in 2003, thus slowing the growth in capital deepening relative to that in the late 1990s. Consequently, the contribution of capital per worker to labor productivity growth fell by 0.7 percentage points between 2001 and 2005 relative to the 1996-2001 period. At the same time, however, TFP growth was accelerating sharply, especially in 2003. The CBO estimates that TFP was responsible for all

Table 3

Contributions of Capital Deepening and TFP to Labor Productivity Growth (1990-2006)

	Average annual growth rate			Change	
	1991-95	1996-2001	2002-06	1991-95 to 1996-2001	1996-2001 to 2002-06
Contribution of capital deepening (percentage points)	0.50	1.33	0.62	0.83	–0.72
Contribution of TFP growth (percentage points)	1.04	1.21	2.07	0.18	0.86
Labor productivity (%)	1.54	2.54	2.65	1.00	0.11

NOTE: Numbers in the table may not add up to totals because of rounding.

SOURCE: CBO using data from the BLS and Bureau of Economic Analysis.

of the acceleration in labor productivity in the 2001-06 period.

A natural question is whether labor productivity will grow as rapidly over the next 10 years as during the past decade. But the experience since 1995 illustrates why that question is so hard to answer. Labor productivity growth is volatile, its measurement is subject to large revisions, and the reasons for changes in its rate of growth are not well understood. Consequently, it is a difficult variable to forecast; past patterns and recent data provide only a rough guide to future labor productivity. Explanations for the recent acceleration help to determine whether any of the changes to growth since 1995 will reverse or recur in the next 10 years.

Projecting Labor Productivity III: Explaining the Acceleration. Although it is hard to say conclusively that one factor is the sole cause of the post-1995 acceleration in productivity growth, most economists point to IT as the primary source. This case is easiest for the late 1990s and more difficult for the period since 2001. As noted previously, the majority of the productivity acceleration for 1996-2000 can be attributed to capital deepening, which was one result of a huge increase in business investment. During the late 1990s, not only did investment boom, but it was heavily tilted toward IT capital (Figure 9). The CBO estimates suggest that faster capital deepening accounted for 80 percent of the upswing in labor productivity growth during the late 1990s and that IT capital accounted for 75 percent of the contribution from capital deepening.

In addition, it appears that rapid technological change in IT industries (including computers, software, and telecommunications) caused faster TFP growth in those industries. It also appears that the pace of technological change was fast enough, and those industries were large enough, for faster TFP growth in that sector of the economy to support overall TFP growth during the late 1990s.[11] However, because TFP growth did not accelerate during the late 1990s, it appears that faster TFP growth in the IT sectors merely offset slower TFP growth elsewhere.

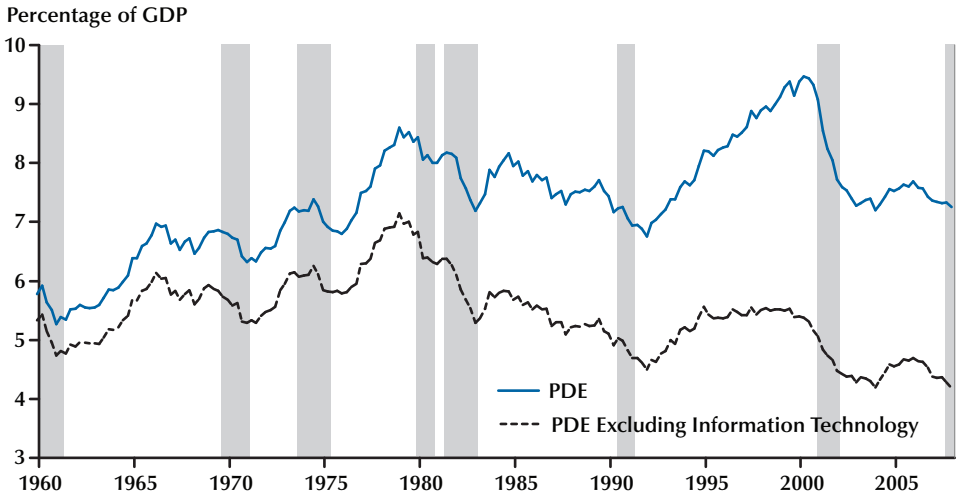
It is somewhat harder to make the case that IT spending was the primary source of the continued rapid growth in labor productivity since the business cycle peak in 2001. One obvious problem with this explanation is that spending on IT capital collapsed after 2000, which strongly suggests that IT capital was not the reason for the continued surge. According to our estimates, nearly 80 percent of the post-2001 growth in labor productivity can be attributed to TFP, with only 20 percent accounted for by capital deepening.

Despite those estimates, the continued growth in labor productivity could still be the result of

[11] According to estimates by Oliner and Sichel (2000), for example, the computer and semiconductor industries accounted for about half of TFP growth from 1996 through 1999, even though those industries composed only about 2.5 percent of GDP in the NFB sector during those years.

Figure 9

Investment in Producers' Durable Equipment



IT spending if a lag exists between the time when the capital is installed and when businesses achieve productivity gains. Several theories, not necessarily mutually exclusive, have been proposed to explain why such a delay could occur. They include the possibility that there are adjustment costs associated with large changes in the capital stock; the possibility that computers are an example of a general-purpose technology, like dynamos and electric motors, that fundamentally change the way businesses operate but take time to produce results; and the possibility that there is a link between IT spending and investment in intangible capital, which refers to spending that is intended to increase future output more than current production but does not result in ownership of a physical asset. As computing power becomes cheaper and more pervasive, managers can invent new business processes, work practices, and organizational structures, which in turn allow companies to produce entirely new goods and services or to improve their existing products' convenience, quality, or variety.

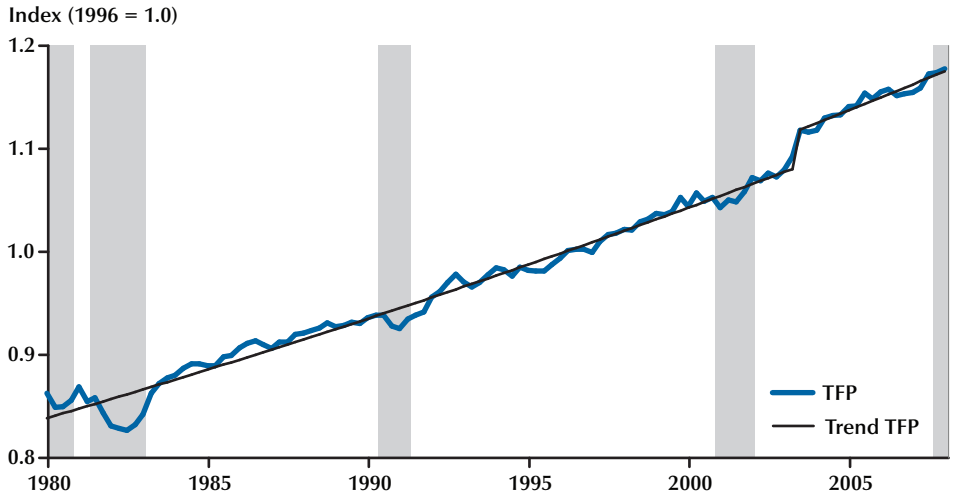
All of these theories could explain the increase in TFP growth. However, all would be expected to have a gradual effect on TFP, raising the growth rate by a small amount over an extended period. In fact, the TFP data display a very steady trend during the 1980s, 1990s, and early 2000s; then a very abrupt increase, occurring entirely in 2003; and then a return to the previous growth trend thereafter (Figure 10). This behavior is somewhat puzzling and hard to reconcile with explanations that rely on a lagged impact of IT spending during the late 1990s. We interpret the abrupt increase as a one-time boost to productivity engendered by the IT revolution—the burst of investment in IT capital allowed firms to raise their efficiency to a higher level but not to permanently increase the rate of productivity growth. Our estimate of potential TFP includes an adjustment that temporarily raises its growth rate to include a level shift similar to that shown in Figure 10.

CONCLUSION

Potential output is a difficult variable to estimate largely because it is an unobservable concept. There are many ways to compute the economy's productive potential. Some methods rely on purely statistical techniques. Others, including the CBO method, rely on statistical procedures

Figure 10

TFP and Trend (1980-2008)



NOTE: Data are adjusted to exclude the effects of methodological changes in the measurement of prices.

grounded in economic theory. However, all of the methods have difficulty estimating the trend in GDP near the end of the data sample, which is usually the period of greatest interest. Because the trend at the end of the data sample is the trend that is projected into the future, any errors in estimating the end-of-sample trend will be carried forward into the projection. The process is further complicated by factors that alter the interpretation of recent economic events, including data revisions and structural change.

In addition to describing the CBO method and highlighting the pros and cons of our approach, this paper describes how we dealt with some developments during the past several years that complicated estimation of potential output. As a general principle, we try to make our estimate of potential output as objective as possible, but as this review of recent problems indicates, estimating potential GDP in real time often involves weighing contradictory evidence. Deciding whether or not, or how much, to change a trend growth rate for TFP, for example, often has a large effect on the estimate of potential for the medium term.

This review demonstrates that the economic landscape is continually changing and that estimates of the trend in any variable, including potential GDP, are affected by those changes. Oftentimes, what looks like a new trend in a series disappears after successive revisions. This factor argues for a conservative approach to estimating such trends and being judicious about changes in those trends.

REFERENCES

Arnold, Robert. "Reestimating the Phillips Curve and the NAIRU." Working Paper 2008-06, Congressional Budget Office, August 2008; www.cbo.gov/ftpdocs/95xx/doc9515/2008-06.pdf.

Brauer, David. "The Natural Rate of Unemployment." Working Paper 2007-06, Congressional Budget Office, April 2007; www.cbo.gov/ftpdocs/80xx/doc8008/2007-06.pdf.

Congressional Budget Office. *CBO's Method for Estimating Potential Output: An Update*. Washington, DC: Government Printing Office,

August 2001; www.cbo.gov/ftpdocs/30xx/doc3020/PotentialOutput.pdf.

Congressional Budget Office. "CBO's Projections of the Labor Force." CBO Background Paper, September 2004; www.cbo.gov/ftpdocs/58xx/doc5803/09-15-LaborForce.pdf.

Congressional Budget Office. *The Budget and Economic Outlook: An Update*. Washington, DC: Government Printing Office, September 2008; www.cbo.gov/ftpdocs/97xx/doc9706/09-08-Update.pdf.

Okun, Arthur M. "Potential GNP: Its Measurement and Significance," in *The Political Economy of Prosperity*. Appendix. Washington, DC: Brookings Institution, 1970; pp. 132-45.

Oliner, Steven and Sichel, Daniel. "The Resurgence of Growth in the Late 1990s: Is Information Technology the Story?" *Journal of Economic Perspectives*, Fall 2000, *14*(4), pp. 3-22.

Stock, James and Watson, Mark. "Variable Trends in Economic Time Series." *Journal of Economic Perspectives*, Summer 1988, *2*(3), pp. 147-74.

Commentary

Robert J. Tetlow

Robert Arnold (2009) clearly and completely lays out the approach used by the Congressional Budget Office (CBO) for measuring potential output and discusses the limitations therein.

In this commentary, I revisit arguments made by the authors and discussants of a paper on this same subject at a 1978 Carnegie-Rochester conference to show how little the CBO methodology differs from methods used 30 years ago and conjecture on why this is so. From there I speculate on why current methods have been impervious to the critiques from 30 years ago and econometric developments in the years thereafter.

The measurement of potential output clearly matters, and matters even more in real time, at least for some decisionmakers. The *growth rate* of potential pins down the tax base for fiscal authorities and lawmakers; it provides a baseline for GDP growth for economic forecasters; and it helps establish a benchmark for policymakers and financial market participants to interpret the real-time data.[1] The *level* of potential defines the point to which the economy is expected to gravitate over the medium term and so is important for monetary authorities, forecasters, and anyone who needs to interpret business cycles. I review why and for whom it matters and critique the methods used by the CBO. The CBO methodology is not unique to that institution; rather, it is my impression that a number of other, large macroeconomic forecast teams around the world use broadly similar tools. To the extent this is true, this critique is germane to a broader set of model builders and users.

After I provide some background, my comments get more specific. I argue that issues of econometric identification limit the confidence with which we can approach the CBO estimates; I argue against the widespread use of deterministic time trends, particularly in the real-time context; and I question the uncritical application of Okun's law.

WHITHER POTENTIAL?

Who needs potential output measures and for what reason? One way of illustrating this question from the perspective of a policymaker is to refer to a simple forecast-based Taylor rule, like the one shown below:

$$(1)\quad R_t = \overbrace{rr^*}^{\partial rr^*/\partial \Delta y^* \geq 1} + \phi_\pi \cdot \overbrace{E_t \pi_{t+1}}^{\partial E\pi/\partial y^* < 0} + \phi_y \cdot \overbrace{\left(y_t - y_t^*\right)}^{\partial\left(y-y^*\right)/\partial y^* < 0} + u_t,$$

[1] An example of the latter is the recent decline in labor force participation. Until a few years ago, sustainable employment growth from the establishment survey was estimated at around 120,000 and levels below that would have been interpreted as foreshadowing possible easing in monetary policy and increases in bond prices. The work of Aaronson et al. (2006) showed that sustainable additions to employment are probably much lower now than before.

Robert J. Tetlow is a senior economist in the Division of Research and Statistics at the Federal Reserve Board. The original discussion slides from the conference are available at the online version of this *Review* article. These slides—but not this text—use Federal Reserve Board/U.S. model vintages and associated databases and *Greenbook* forecast and historical databases (the latter of which under Federal Reserve Board rules are permissible for use only for forecasts currently dated before December 2002). The author thanks Peter Tulip, Dave Reifscheider, and Joyce Zickler for useful comments and Trevor Davis for help with the presentation slides.

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 291-96.
© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

where R is the nominal federal funds rate, rr is the real funds rate, π is inflation, y is (the natural logarithm of) real gross domestic product (GDP), and u is a stochastic term. The asterisks on the real rate and on output represent "potential" (or "natural") levels; these natural levels are not observable. The coefficients, ϕ_j, $j = y$, and π, would normally be expected to be positive. The partial derivatives above the equation itself show how changes in potential output affect the rule and hence decisionmaking. Starting with the term farthest to the right, an increase in the level of potential—that is, $\partial y^* > 0$—decreases estimates of the output gap, $y - y^*$, all else equal. Higher potential would also reduce expected future inflation—$E_t\pi_{t+1}$—because smaller gaps usually mean less inflation and both of these would be expected to lead to a lower federal funds rate. An increase in the growth rate of potential—$\partial(\Delta y^*) > 0$—raises the equilibrium real interest rate, which would call for an increase in the funds rate, all else equal, but it would also have complex, model-dependent effects on current and future output gaps and inflation.[2]

What complicates this is that the only observables in the equation are current output, which is subject to revision, and the federal funds rate itself. A policymaker—in this instance, the Fed—is obliged to add structure to this underidentified equation through the use of a macroeconomic model of some sort. For their part, interpreters of the data—Fed watchers, among others—are obliged to "invert" the (perceived) policy rule and infer what the Fed's estimates of rr^*, Δy^*, y^*, and $E_t\pi_{t+1}$ might be.[3] The only inevitability is that *all* parties will get it wrong; the question is in what way and how critically.[4]

[2] I am thinking of a closed economy here, or at least one that, if open, is not "small."

[3] Of course, what Fed watchers might also want to infer from policy decisions given a policy rule is an estimate of the target rate of inflation. The target rate has been normalized out of our policy rule, for simplicity.

[4] It makes a difference whether it is the Fed that is "getting it wrong" or the private sector. The more the Fed gets things wrong, the harder it is for the private sector to infer something about the economy from Fed behavior. This is, of course, one of the reasons behind arguments for transparency in monetary policy.

METHODOLOGY: A DÉJÀ VU EXPERIENCE

Bob Arnold's paper does a solid job of explaining the CBO's methodology for measuring and projecting potential output. He also shows substantial awareness of the limitations of their approach; there is little for me to add on that score. To provide a different perspective, in this section I offer readers a "blast from the past," from 30 years ago, in fact. I describe the approach of Perloff and Wachter (1979) from a Carnegie-Rochester conference in 1978. Like Arnold, Perloff and Wachter start with an estimate of the non-accelerating inflation rate of unemployment (NAIRU) from a previous paper; then, they estimate potential labor input as follows:

$$(2)\quad \log(n) = c + \alpha\left(u - u^*\right) + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \varepsilon_t,$$

where t^k, $k = 1,2,3$ are polynomial time trends, u is the unemployment rate, c is a constant, and ε is a residual. Potential labor input, n^*, is evaluated using this equation by setting cyclical and noise terms to zero; in this instance, $u = u^*$ and $\varepsilon = 0$ for all t. Perloff and Wachter follow the same procedure with potential capital input, except that the equation in this case is a "cyclically sensitive translog production function" (p. 122) augmented with more polynomial time trends. The similarity to Arnold's equation (1) is remarkable.[5]

With this sameness in mind, I can make my job as discussant easier by shamelessly stealing from Perloff and Wachter's discussants. Gordon (1979) focused on estimation:

> [W]ithout making any statistically significant difference in the wage equation, one could come up with an estimated increase in u^* between 1956 and 1974 ranging anywhere from 0.58 to 1.61 percentage points… (p. 190)

In other words, taking u^* as exogenous, rather than estimating a complete system, particularly while ignoring the imprecision of the first-stage

[5] From a real-time perspective, the CBO's methodology could be more problematic than Perloff and Wachter's in that the CBO uses trends dated back from the previous business cycle peak. No doubt this is to avoid the political heat that might come from making a call on a potentially contentious issue in real time. By definition, this method will miss turning points, possibly by wide margins.

estimates, is problematic. Elsewhere, Gordon remarks on overparameterization:

> Taking this set of data for u^*, one can compute an acceptable and consistent natural output series without any use of production functions at all. (p. 188)

That is, because the time-trend variables are doing the bulk of the work, it is not clear that there is anything unambiguously "supply side" in the calculation. The other discussants, Plosser and Schwert (1979), focused on interpretation of the results and the related issue of econometric identification:

> [A]ggregate demand policies are not necessarily appropriate in a world where actual output is viewed as the outcome of aggregate supply and demand...In such an equilibrium world, "potential output" ceases to have any significance. (p. 184)

Thus, even though the real business cycle literature had yet to emerge, the seeds of the idea were clearly already planted.

Both commentaries remark, in their own way, on econometric identification. How does one differentiate between supply (or potential) and demand (the gap)? Does it even make sense to try? The use of time trends, which are both *deterministic* and *smooth*, is an identifying assumption made by both Perloff and Wachter (1979) and Arnold (2009). Their use implies that supply shocks have not happened often historically and can be safely ignored in real time for forecast purposes. When Perloff and Wachter were writing, the literature on unit roots in real GDP—which would come to include, as it happens, an important contribution by Nelson and Plosser (1982)—had not yet arisen. But this is not so for the CBO or any of a variety of other institutions that use similar approaches.[6] Why, then, has the methodology on measuring potential output apparently not absorbed anything from the literature on unit roots and stochastic trends over the past 30 years?

My conjecture is threefold. First, the CBO—like most macroeconomic policy institutions—maintains a distinctly Keynesian perspective on how the economy works, a view that maintains that the majority of fluctuations in real GDP come from demand disturbances and that policy plays a key role in smoothing those fluctuations. This approach is natural enough; policy institutions do tend to draw individuals who believe that policy is highly consequential. And to paraphrase the old line: When one likes to use hammers, the object of interest tends to look like a nail. My second conjecture is more subtle. Economists at institutions like the CBO must be able to answer a wide variety of questions from decisionmakers and they need a structure that allows them to do so in short order. The complex, deterministic accounting structure that Arnold describes allows the CBO to do that, although one could of course quarrel with the efficacy of the advice that comes from such a structure. Third, while I would argue that the literature on unit roots shows that permanent shocks to GDP—shocks that can fairly be characterized as supply shocks—are important, that literature has not yet provided high-precision tools for measuring those shocks in real time. The standard errors of estimates of potential output and the output gap are large.[7] And the problem gets worse as the parameter space of the model grows.

Nonetheless, I would argue that even though adopting the stochastic approach involves tackling some difficult issues, it is still a step worth taking. These same issues exist with the extant method, but they have been swept under the rug through the identification by assumption implicit in the use of time trends to represent aggregate supply. We are dealing with unobserved variables here; it only makes sense that, with the passage of time, our backcasts of potential output would differ significantly from our nowcasts. To "assume away" the stochastic properties of the data only

[6] Barnett, Kozicki, and Petrinec (2009) note that the Bank of Canada has used a stochastic method for measuring potential since 1992. The Federal Reserve Board's FRB/U.S. model forecast uses a stochastic state-space method. The Fed's official *Greenbook* forecast—being judgmental—is more complicated. The Board staff consult a variety of models for guidance on adjusting potential output and its constituent parts, but they do so on an ad hoc basis. There is, however, a significant smoothness prior on trend labor productivity, and hence on potential output, and a prior that Okun's law holds fairly strongly.

[7] The discussion slides show an example of the bootstrapped standard errors from a simple unobserved components model of potential output. These are available at the online version of this *Review* article.

ignores the issue; it doesn't solve it. A more clear-eyed view, in my opinion, is to accept the stochastic nature of potential and adjust procedures and interpretations to this reality by being prepared to adapt estimates rapidly and efficiently in real time (see, e.g., Laxton and Tetlow, 1992).

OKUN'S LAW

I have already noted the strong Keynesian prior implicit in the methods for measuring potential output at the CBO and other policy institutions. As noted, this prior is evident in the use of deterministic time trends. It is also a function of the fact that potential output—and hence output gaps—are constructed beginning with estimates of the NAIRU, and hence the unemployment gap, using Okun's law. This is illustrated in Arnold's Figure 1, which shows the CBO output gap and the unemployment gap on the same chart. The chart provides an "ocular regression" of Okun's law: The two lines are nearly on top of one another, meaning that a linear, static relationship between the two concepts fits the (constructed) data very well. In essence, this means that the output gap and the unemployment gap are nearly the same.

The view that the unemployment gap and output gap are isomorphic—that is, the view that Okun's law really is something that approaches a "law"—has important implications for the characterization of business cycles. The following log-linearized Cobb-Douglas production function shows this:

$$y = a + \theta n + (1-\theta)k, \tag{3}$$

where a is total factor productivity, and we measure potential output using full-employment labor input, n^*, and the actual capital stock, k, as is usually the case:

$$y^* = a^* + \theta n^* + (1-\theta)k, \tag{4}$$

and then subtract equation (4) from equation (3) to show the relationship between output gaps, $y - y^*$, and the labor market gap, $n - n^*$[8]:

$$y - y^* = (a - a^*) + \theta(n - n^*). \tag{5}$$

Now Arnold's Figure 1 implies that $y - y^* - \theta(n - n^*)$ is small and unimportant—taken to the limit, Okun's law implies that it should be white noise. This, in turn, means that what we might call the *productivity gap*, $a - a^*$, must also be small and unimportant. Should it be? Should anyone care? What is the productivity gap anyway? The productivity gap can represent any or all of a variable workweek of capital, variable capacity utilization, or labor adjustment costs to productivity shocks.[9] Loosely speaking, fluctuations in a that are not in response to shocks to a^* are labor adjustment shocks, whereas shocks to a^*, all else equal, are classic productivity shocks. The productivity gap, $(a - a^*)$, can be unimportant only in the unlikely circumstance that actual productivity, a, moves instantaneously with a productivity shock, a^*, and disturbances to a, holding a^* constant, are themselves close to white noise. In short, the only way the productivity gap could be small and unimportant—and, therefore, the only way that Okun's law can hold so tightly as to be called a law—is either because aggregate demand moves instantaneously with productivity or if there are no productivity shocks in the first place. Neither of these possibilities seems plausible.

My own preference would be to drop the deterministic time trends, relaxing somewhat the iron grip of Okun's law, and treat potential output as a stochastic variable. Doing so would allow for meaningful supply-side shocks, modeled using state-space techniques, probably with the Kalman filter.[10] From an operational point of view, this shifts the prior on the incidence of shocks somewhat. Under the deterministic prior, all real surprises are demand shocks and this view

[8] I am blurring the distinction between the unemployment gap and the labor market gap—the difference being what might be called the average workweek gap and the labor force participation rate gap. This distinction is important to my point only if one thinks that *all* productivity adjustment—a movements relative to a^*—is carried out on these two margins, which seems unlikely.

[9] Whether there is any meaningful distinction among these three stories depends on the underlying model.

[10] My suggestions here are particularly relevant for a decisionmaking body when the level of the gap is important. I think this is true for almost all policy institutions but is undoubtedly "more true" for, say, a central bank, than for a fiscal authority.

is adjusted only rarely and after the fact; with the stochastic view, the default option becomes one wherein some portion of a given output surprise is characterized as a supply shock. The model user could override that prior, but it would be a conscious decision on the user's part to do so. In this way, the stochastic approach would be responsive in real time, allowing estimates to adapt to developments such as the productivity boom of the late 1990s in a way that the deterministic approach would not. Such a property is an important one, particularly for institutions whose policy instruments may be adjusted with relatively high frequency. State-space models also allow the modeling of nonlinearities—for example, to capture different dynamics when cycles are being driven largely by supply shocks rather than by demand shocks or to allow for "jobless recoveries"—although the econometric hurdles are correspondingly higher.[11]

Such an approach comes at some cost, however, because either the parameter space must be small or the user must be willing to impose priors on enough parameters to give the estimator a chance of producing reasonable results. Still, this approach would likely impose fewer restrictions than the current approach. At a minimum, weakening the prior that all shocks are demand shocks opens the door for model users to consider what kind of shocks might have produced the cross section of measured surprises—positive for output and negative for inflation, for example—in real time. This, in turn, would allow a more rapid adjustment to new information and smaller and less persistent forecast and policy errors than would otherwise be the case.

CONCLUSION

Bob Arnold has outlined a detailed and sophisticated approach to measuring potential output as used by the CBO. In my opinion, the approach is representative of the perspective and needs of a range of policy institutions. In general terms, the remarkable thing about the CBO method, and methods like it, is how little it differs from methods used 30 years ago. This lack of penetration of academic ideas into the policymaking sphere is perplexing in some ways. However, it reflects, in part, the needs of institutions to be able to answer myriad questions using the same model. This practice tends to result in the construction of large, elaborate models, and unfortunately not all modern econometric techniques scale up well to large models. The good news is that new methods in Bayesian econometrics offer considerable help in estimating larger systems while paying proper heed to the priors of the model builders and users. Another source of the lack of progress, in my view, is the strong Keynesian prior regarding the sources of business cycle fluctuations. Many public policy institutions regard supply shocks as rare enough to be ignored. I would argue that this prior is overly strong—we know for a fact it was dead wrong in the United States in the late 1990s (see, e.g., Anderson and Kliesen, 2005; and Tetlow and Ironside, 2007). It might also be deleterious for policymaking because the perspective that all shocks are demand shocks leads directly to the view that all fluctuations should be smoothed out, which is arguably a recipe for "fine-tuning."

We are now in a period in which the CBO methodology is being tested. By construction, the CBO will have concluded that the current "financial stress shock" to the U.S. economy is entirely a demand-side phenomenon with large implications for the output gap and eventually for inflation. This is a contestable position. It would not be hard to fashion an argument that the desired capital stock, and hence the level of potential output, has shifted down; interpreting the shock in this less devoutly Keynesian way would mean smaller output gaps, less disinflationary pressure, and somewhat less need for expansionary policy, all else equal. We shall see. In any case, quite apart from the methods detailed therein, Bob Arnold's paper shows a mindful understanding of the uncertainties involved, which is probably more important. It thereby serves the Congress well.

[11] Bayesian methods can be helpful in this regard, particularly for policy institutions that tend to be unapologetic about having prior beliefs.

REFERENCES

Aaronson, Stephanie; Fallick, Bruce; Figura, Andrew; Pingle, Jonathan and Wascher, William. "The Recent Decline in the Labor Force Participation Rate and Its Implications for Potential Labor Supply." *Brookings Papers on Economic Activity*, Spring 2006, *1*, pp. 69-134.

Anderson, Richard G. and Kliesen, Kevin L. "Productivity Measurement and Monetary Policymaking During the 1990s." Working Paper No. 2005-067A, Federal Reserve Bank of St. Louis, October 2005; http://research.stlouisfed.org/wp/2005/2005-067.pdf.

Arnold, Robert. "The Challenges of Estimating Potential Output in Real Time." Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 271-90.

Barnett, Russell; Kozicki, Sharon and Petrinec, Christopher. "Parsing Shocks: Real-Time Revisions to Gap and Growth Projections for Canada." Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 247-65.

Gordon, Robert. "A Comment on the Perloff and Wachter Paper." *Carnegie-Rochester Conference Series on Public Policy*, January 1979, *10*(1), pp. 187-94.

Laxton, Douglas and Tetlow, Robert J. "A Simple Multivariate Filter for the Measurement of Potential Output." Technical Report No. 59, Bank of Canada, June 1992.

Nelson, Charles R. and Plosser, Charles I. "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications." *Journal of Monetary Economics,* September 1982, *10*(2), pp. 139-62.

Perloff, Jeffrey M. and Wachter, Michael L. "A Production Function—Nonaccelerating Inflation Approach to Potential Output: Is Measured Potential Output Too High?" *Carnegie-Rochester Conference Series on Public Policy*, January 1979, *10*(1), pp. 113-63.

Plosser, Charles I. and Schwert, William G. "Potential GNP: Its Measurement and Significance: A Dissenting Opinion." *Carnegie-Rochester Conference Series on Public Policy*, January 1979, *10*(1), pp. 179-86.

Tetlow, Robert J. and Ironside, Brian. "Real-Time Model Uncertainty in the United States: The Fed, 1996-2003." *Journal of Money, Credit and Banking*, October 2007, *39*(7) pp. 1533-61.

Trends in the Aggregate Labor Force

Kenneth J. Matheny

Trend growth in the labor force is a key determinant of trends in employment and gross domestic product (GDP). Forecasts by Macroeconomic Advisers (MA) have long anticipated a marked slowing in trend growth of the labor force that would contribute to a slowing in potential GDP growth. This is reflected in MA's forecast that the aggregate rate of labor force participation will trend down, especially after 2010, largely in response to the aging of the baby boom generation, whose members are beginning to approach typical retirement ages. Expectations for a downward trajectory for the participation rate and a slowing in trend labor force growth are not unique. However, this article reports on MA research suggesting that the opposite is possible: that the slowdown in trend labor force growth could be relatively modest and that the trend in the aggregate rate of labor force participation will decline little, if at all, over the next decade. (JEL E01, J11)

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 297-309.

Projections of population and labor force growth are essential elements of any projection of the economy's potential output growth. Often, however, these projections are driven primarily by trends and dummy variables. The research reported here constructs a labor force projection from a much richer set of behavioral determinants of labor force trends than are typically used. The set of determinants also is richer than that contained in the aggregate labor force equation that appears in the current version (as of this writing) of the Macroeconomic Advisers (MA) commercial macroeconomic model.

In its September 24, 2008, issue of *Long-Term Economic Outlook*, MA projected that the labor force participation rate would decline by about 1½ percentage points over the next decade—to 64.6 percent in 2017—and that the growth of the labor force would slow from roughly 1 percent or a little higher on average in recent years to an average of 0.6 percent from 2013 to 2017 (Tables 1 and 2). These estimates are comparable to recent estimates from the Congressional Budget Office. However, they are considerably stronger than trend estimates in a recent paper by Aaronson et al. (2006). Their research suggests that demographic and other developments could result in a much larger decline in the participation rate—to 62.5 percent by the middle of the next decade—and a reduction in trend labor force growth to just 0.2 percent from 2013 to 2015.

The research summarized here leans in the other direction. It suggests that trend growth of

Kenneth J. Matheny is a senior economist at Macroeconomic Advisers, LLC. The author thanks James Morley of Washington University for research advice and assistance. Other staff at Macroeconomic Advisers contributed to this research in various ways, including Joel Prakken, chairman; Chris Varvares, president; Ben Herzon, senior economist; Neal Ghosh, economic analyst; and Kristin Krapja, economic analyst. The author also acknowledges the following for their assistance or feedback: Robert Arnold, Congressional Budget Office; Jonathan Pingle, Brevan Howard Asset Management, LLP; Mary Bowler, Sharon Cohany, John Glaser, Emy Sok, Shawn Sprague, and especially Steve Hipple and Mitra Toosi of the Bureau of Labor Statistics; Steven Braun of the President's Council of Economic Advisers; and William Wascher of the Federal Reserve Board of Governors. The staff of Haver Analytics provided assistance locating certain data. Ross Andrese, a former employee of Macroeconomic Advisers, provided research assistance during an early phase of this project.

© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

Table 1

Growth of the Civilian Labor Force

Year	MA *Long-Term Economic Outlook* (2008)	Model prediction of trend*	CBO (2008) estimate of trend	Aaronson et al. (2006) estimate of trend
2008	0.8	1.3	1.1	0.4
2009	0.8	1.1	1.0	0.4
2010	1.1	0.9	0.9	0.4
2011	1.0	0.9	0.6	0.4
2012	0.8	0.9	0.6	0.3
2013	0.6	1.0	0.6	0.2
2014	0.6	0.9	0.5	0.2
2015	0.6	0.9	0.5	0.2
2016	0.6	0.9	0.5	NA
2017	0.6	0.9	0.4	NA

NOTE: Data represent annual averages in percent. *Based on the level terms of the regression in Table 3 after removing cyclical contributions from the unemployment and wealth terms, as described in the text.

Table 2

Labor Force Participation Rate

Year	MA *Long-Term Economic Outlook* (2008)	Model prediction of trend*	CBO (2008) estimate of trend	Aaronson et al. (2006) estimate of trend
2008	66.0	65.7	66.1	65.2
2009	65.9	65.8	66.0	64.7
2010	65.8	65.7	65.9	64.4
2011	65.8	65.7	65.7	64.0
2012	65.7	65.7	65.4	63.7
2013	65.5	56.8	65.2	63.3
2014	65.3	65.9	64.9	62.9
2015	65.1	66.0	64.6	62.5
2016	64.9	66.0	64.3	NA
2017	64.6	66.0	63.9	NA

NOTE: Data represent annual averages in percent. *Based on the level terms of the regression in Table 3 after removing cyclical contributions from the unemployment and wealth terms, as described in the text.

the labor force might not slow as much over the next decade as previously anticipated. According to the model, the trend in labor force growth will edge down slightly to an average of 0.9 percent through 2017, and the trend in the labor force participation rate will dip only slightly from recent levels to average just under 66 percent from now through 2017. The research reported here updates our measure of the pure demographic contribution to the change in the labor force to reflect more age detail than in our existing model and to match the population concept on which it is based with the one that underpins the official estimates of the labor force and the participation rate from the Bureau of Labor Statistics (BLS). The updated model addresses a bedeviling problem with discontinuities in the official estimates of the labor force and the civilian noninstitutional population. Unfortunately, data limitations prevent the complete elimination of the spurious impacts of these discontinuities, which stem from updates to "population controls" that are entered into the official data in response to the results of decennial censuses and for other population-related data.

The research reported here shows a much richer set of behavioral determinants of labor force trends than are contained in the equation for the aggregate labor force that appears in the MA commercial model at the time of this writing. Specifically, this analysis drops previously used deterministic trend and shift terms; instead, the model includes a small set of factors believed to exert important behavioral influences on the labor force.

DEMOGRAPHIC CONTRIBUTION TO THE LABOR FORCE

As part of our modeling, the pure demographic contribution to the change in the labor force is separated from its behavioral influences. We typically measure the demographic contribution with a chain-weighted index of the populations for 30 different age and gender brackets, using lagged labor force participation rates as weights, which we label *LFCADJL*.[1] Population details from the civilian noninstitutional population 16 years and older are used to construct the series. With lower-case *p*'s denoting participation rates and lower-case *nc*'s denoting population details, *LFCADJL* is updated according to

$$LFCADJL_t = LFCADJL_{t-1} \times \left(\sum_{i=1}^{30} p_{i,t-1} \times nc_{i,t}\right) \Big/ \left(\sum_{i=1}^{30} p_{i,t-1} \times nc_{i,t-1}\right). \tag{1}$$

The series is indexed to equal the actual labor force in 2000. (This has no impact on the results that follow.) Changes in *LFCADJL* from one quarter to the next are due to changes in the detailed populations across age and gender brackets (the *nc*'s) holding fixed the weights (the *p*'s). In this sense, growth of *LFCADJL* is a comprehensive measure of the pure demographic contribution to the change in the labor force. Growth of the actual labor force and growth of *LFCADJL* are displayed together in Figure 1.[2] Forecast projections for *LFCADJL* reflect growth in the population detail, holding fixed the within-group participation rates. We observe that growth of *LFCADJL* is projected to moderate in the forecast, with an average of 0.4 percent from 2015 to 2017.

BEHAVIORAL COMPONENT OF THE LABOR FORCE

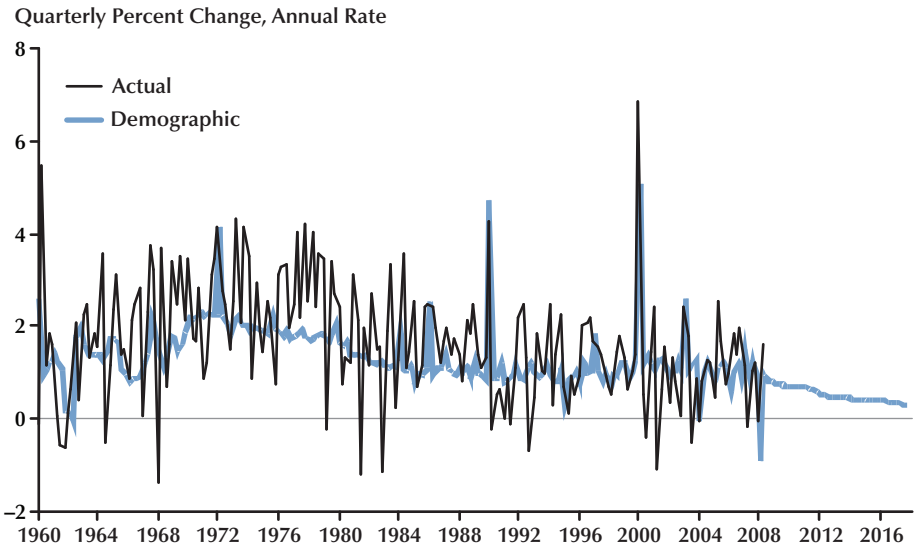
The behavioral component of the labor force can be measured by the log-ratio of the actual labor force to the demographic measure, log(*LFC*/*LFCADJL*). This series (Figure 2) is obviously nonstationary, and tests confirm that it appears to be I(1), that is, the series is stationary after differencing, implying that cointegration-based techniques provide a useful framework for econometric analysis. We found evidence that this

[1] Male and female populations and labor forces are separated into 15 non-overlapping age brackets, specifically, 16-17, 18-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-61, 62-64, 65-69, 70-74, and 75 years and older.

[2] We return in a subsequent section to the appearance of sharp swings in the growth rates of these series stemming from updated population controls that are entered into the population and labor force data without adjustment. The most recent discontinuity occurs in data for the first quarter of 2008, reflected in a sharp, temporary drop in the growth of *LFCADJL*.

Figure 1

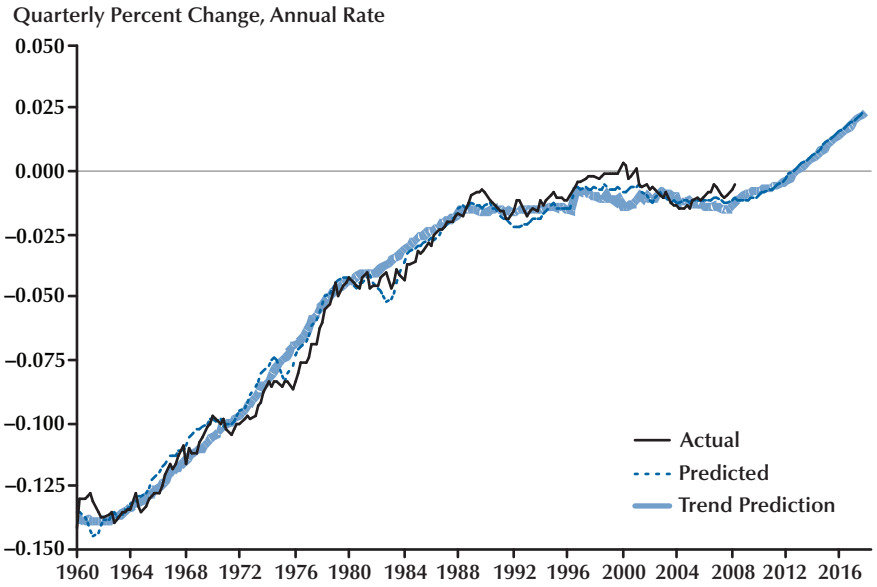
Labor Force Growth: Actual and Demographic Contribution



SOURCE: BLS and MA.

Figure 2

Prediction for Behavioral Component



variable is cointegrated with the following set of "behavioral" variables (and a constant).

Dependency Ratio (*YOUNG015*): The ratio of persons 15 years and younger to the entire resident population. This series has generally trended down over the past several decades, roughly mirroring the inverse of the relative participation rate for women. This term is typically among the most robust and statistically significant variables in labor force regressions.

Life Expectancy (*WT65F_LEF65*): The life expectancy of women at the age of 65 years, multiplied by the share of women aged 65 and older in the total adult, civilian, noninstitutional population. Life expectancy represents the number of years one would expect to live, on average, conditional on having attained the age of 65.[3] A subsequent section addresses our choice of a female-weighted, female life expectancy.

Welfare Reform (*WR1996*): Intended as a proxy for the effect of welfare reform in the late 1990s. This series is constructed as the product of several terms, beginning with a dummy variable that is zero up to the second quarter of 1996 and one thereafter, to mark the enactment of federal welfare reform in August 1996.[4] The zero-one dummy is multiplied by one minus the share of women who are married, by the dependency ratio (*YOUNG015*), and by the ratio of the population of women aged 18 to 49 to the total adult civilian noninstitutional population.[5]

Household Net Worth (*NW_SCALED5564*): The ratio of per capita household net worth to hourly labor compensation, multiplied by the population share of persons aged 55 to 64. The traditional theory of the labor/leisure choice notes that increases in wealth cause a reduction in labor supply if leisure is a "normal" good. However, previous research on the existence of wealth effects on labor supply has been mixed.[6] We found ambiguous results when the wealth ratio is not scaled by the population share but robust results consistent with traditional theory when the wealth ratio is premultiplied by the share of the population aged 55 to 64. In other research on participation rates for individual age brackets, we found evidence of wealth effects on participation rates for this age bracket.

Unemployment Rate (*LURC*): The official unemployment rate, expressed in percent. Its presence is motivated by search-theoretic considerations, namely, that the expected return to searching for employment is negatively related to the level of unemployment.

A simple levels regression among these variables and a constant suggested cointegration, so a dynamic levels regression was estimated that also includes leads and lags of the first differences of all the regressors to control for serial correlation.[7] The results of the dynamic regression are summarized in Table 3.[8,9] All five regressors enter as expected, with positive coefficients on the life expectancy and welfare reform terms, and negative coefficients on the dependency ratio, the

[3] Estimates for life expectancy are from the "intermediate-cost" assumptions of the Social Security Administration (www.ssa.gov/OACT/TR/TR07/lr5A4.html). Interpolation from annual to quarterly estimates is accomplished using a cubic spline. We take a centered nine-quarter moving average to smooth sometimes odd movements in the first differences that arise because of interpolation. Smoothing has very little effect on the regression results.

[4] The *Personal Responsibility and Work Opportunity Reconciliation Act of 1996* was signed into law by President Clinton on August 22, 1996. Some states began instituting welfare reforms during the same era or before. We also considered slightly different versions of this term where the dummy variable switches from zero to one either before or after the third quarter of 1996. For dates near the third quarter of 1996, the regression results were little affected.

[5] One might suppose that, in the regression, the welfare reform term is capturing a behavioral increase in the labor force as persons were "pulled" into labor markets during a period of strong labor demand beginning in the late 1990s. We discount this possibility for two reasons. First, the welfare reform term is significant with the unemployment rate present, and the unemployment rate arguably accounts for any "demand pull" effect. Second, the size of the effect from the welfare reform term is comparable to estimates from other researchers about the impact of welfare reform in the 1990s.

[6] Goodstein (2008) finds that increases in wealth do lead to earlier withdrawal from the labor force in a panel dataset of older men. He argues that previous researchers who investigated the issue in panel datasets found small and statistically insignificant effects of wealth on retirement because they did not control for differences in "tastes," including risk aversion and preference for work, thereby producing a spurious positive correlation between wealth and labor force participation.

[7] Along with a correction for heteroskedasticity, *t*-statistics from the dynamic levels regression are asymptotically valid.

[8] To conserve space, the differenced terms, which are immaterial to what follows, are suppressed in the table.

[9] Sample means of the first-differenced terms were removed before estimation. This has no effect on the estimated coefficients, except for the constant, and ensures that the predicted value of the level terms is consistent with the level of the dependent variable during the estimation sample.

unemployment rate, and the wealth term. All terms are statistically significant.

We noted at the outset that the primary focus of this research was on the determinants of trends in the labor force. Toward that end, we removed the direct "cyclical" contribution by replacing the unemployment rate (*LURC*) with our estimate of the long-run natural rate of unemployment (NAIRU).[10] The wealth term is also subject to cyclical influences, though the issue of identifying its cyclical contribution is ambiguous. On the one hand, it might not matter much in the forecast beyond 2010, because the contribution from the wealth term does not vary much after that date. Nevertheless, we did attempt to reduce the obvious cyclicality in the wealth term as follows. First, we regressed the unscaled wealth ratio (that is, per capita wealth divided by hourly compensation without scaling by the population share) on several leads and lags of the unemployment rate, along with a constant and trend. We then substituted the contribution from the unemployment rate with a contribution computed using the NAIRU and the same coefficients. The adjusted wealth rate was once again multiplied by the 55- to 64-year-old population share (*NW_SCALED5564LR*).[11]

With these adjustments, the model for "trend" in the behavioral component of the labor force is given by

$$\begin{aligned}(2)\quad & 0.1233 + 0.0784 \times WT65F_LEF65_t - 0.9330 \\ & \times YOUNG015_t + 0.2146 \times WR1996Q3_t - 0.2682 \\ & \times NW_SCALED5564LR_t - 0.0048 \times \mathrm{NAIRU}_t.\end{aligned}$$

The coefficients in this expression are identical to those on the corresponding level terms in Table 3. The predicted value for this model over both history and forecast is displayed in Figure 2, along with a prediction that does not remove cyclical contributions from the unemployment rate. Forecast assumptions for the wealth ratio and the NAIRU are from MA's most recent *Long-term Outlook* publication.

The model easily incorporates the secular increase in the log-ratio from the early 1960s to the 1990s. It also easily replicates the flattening that began in the late 1990s and, to some extent, the downturn in the first half of the current decade. As of 2008:Q2, the actual and predicted ratios differ by just 0.6 percent. According to the model, about three-fourths of the increase in the ratio of *LFC* to *LFCADJL* is "explained" by the dependency term, with most of the remainder accounted for by life expectancy, with smaller and roughly offsetting contributions on net from the other terms. According to the model, welfare reform raised the level of the labor force by approximately 0.75 percent beginning in 1996:Q3, or by about 1.0 million persons. This figure is comparable to estimates by other researchers of the impact of welfare reform.[12] To a first approximation, the impact on the labor force from the welfare reform term is nearly constant through the end of the estimation sample and in the forecast.[13] The esti-

[10] Our estimate of the NAIRU is not a constant because it includes a gradually evolving adjustment for changes in the age profile of the labor force. For example, younger adults on average experience higher unemployment rates, so an increase in their share of the labor force would, all else equal, be associated with an increase in the unemployment rate.

[11] An alternative procedure to reduce the influence of cyclical movements in the unemployment rate on the model's prediction for the labor force would be to replace the unemployment rate with the NAIRU and to replace the original wealth term with the "adjusted" version when estimating the regression. In this alternative, the NAIRU is not statistically significant, but the coefficient on the adjusted wealth term is little changed. Moreover, there is a substantial increase in the coefficient on the life expectancy term that leads to a much higher forecast for the participation rate—approximately 2 percentage points higher by 2017—which we would be uncomfortable showing as a base-case scenario. In any event, this exercise suggests that the forecast projections based on the original model (derived from the level terms in the regression in Table 3) are not overly optimistic.

[12] Blank (2004) notes that between 1995 and 2001, a period over which, on net, there was little change in the aggregate unemployment rate, employment of single mothers rose by approximately 820,000, as welfare caseloads fell by roughly double that amount. The author argues that 820,000 likely understates the full effect on employment of welfare reform. The impact on the labor force was likely even larger than the impact on employment. Bartik (2000) estimated that welfare reform expanded the labor force of less-skilled women by over 1 million persons.

[13] The value of *WR1996* rises from zero to about 0.035 in 1996:Q3. On balance, it drifts down through the end of the estimation sample, to about 0.031 as of 2008:Q2. Based on the estimated model in Table 3, the percentage contribution from this term declined from about 0.75 percent in late 1996 to about 0.66 percent in early 2008. In level terms, the estimated contribution to the labor force in early 2008 (of 1 million persons) is essentially identical to the contribution from this term as of 1996:Q3.

Table 3

Summary of Regression Results
Dependent Variable: log(*LFC/LFCADJL*)
Sample: 1960:Q1–2008:Q2
Included observations: 194

Variable	Coefficient	HAC SE	*t*-Statistic	*p*-Value
CONSTANT	0.0123	0.0392	3.1492	0.0020
YOUNG015	–0.9330	0.0556	–16.7784	0.0000
WTF65F_LEF65	0.0784	0.0135	5.7987	0.0000
WR1996	0.2146	0.0761	2.8191	0.0055
LURC	–0.0048	0.0005	–10.7430	0.0000
NW_SCALED5564	–0.2682	0.0436	–6.1487	0.0000

R^2	0.9968	Mean dependent variable	–0.0524
Adjusted R^2	0.9960	SD dependent variable	0.0473
Standard error of regression	0.0030	Akaike information criterion	–8.5967
Sum squared residual	0.0014	Schwarz criterion	–7.9061
Log likelihood	874.88	*F*-statistic	1,195.81
Durbin-Watson statistic	0.7750	Probability (*F*-statistic)	0.0000

NOTE: HAC SE, heteroskedasticity and autocorrelation consistent standard error; SD, standard deviation. Not shown are the coefficients on the leads and lags of first differences for each of the level regressors (excluding the constant). Three leads and lags and contemporaneous values were included for each of the differenced terms. Sample means were deducted from the first differences before estimation.

mate of "trend" for the behavioral component is a little higher than the unadjusted prediction for periods when the unemployment rate is above the NAIRU.

The model's forecast includes a pronounced upward movement in the behavioral component of the labor force, especially after 2011, mostly in response to an increasing (indeed, accelerating) contribution from the life expectancy term, along with a small increase in the contribution from the dependency term. The contribution from the welfare reform is nearly a constant in the forecast, and the contribution from the adjusted wealth term to the change in the forecast through 2017 is small. We return to a discussion of the life expectancy term and its contribution to the forecast in a subsequent section.

DISCONTINUITIES IN POPULATION CONTROLS

The historical time series on the civilian noninstitutional population periodically exhibits sharp swings stemming from changes in the population controls that are used to extrapolate survey results (population data are published by the U.S. Bureau of the Census in *Current Population Survey*). The reason is that when the population controls are updated, their effects are not normally backdated or smoothed when entered into the official estimates for the civilian noninstitutional population. For example, when the population control for January 2000 was raised to reflect the results of Census 2000, it led to an upward adjustment to the official estimate for the civilian noninstitutional population as of that date of

approximately 2.6 million persons. Data for previous periods were not restated upward to reflect the new, higher population control for January 2000, resulting in a discontinuity in the official data.[14] Similar discontinuities surround previous decennial censuses and other dates. Discontinuities exist for the same reason in the official data on the labor force.

The existence of discontinuities affects our measure of the demographic contribution to the labor force, *LFCADJL*, because the population details used in its construction are subject to the same discontinuities. This does not represent a problem for our regression analysis, because the estimates of the civilian labor force (*LFC*) and *LFCADJL* are subject to discontinuities from the same source and are consistent. However, the existence of population control–related discontinuities does affect estimation of "trend" in these series (and in the civilian noninstitutional population).[15]

Estimates of the effect of revised population controls on the aggregates for the civilian noninstitutional population and for the total labor force are available in BLS publications for several decades of data, but detailed information necessary to smooth the impacts on the population details used to construct *LFCADJL* is not available. Given these discontinuities, what is the best way to proceed? Although highly imperfect, we adjust *LFCADJL* by multiplying it by the ratio of the adjusted to the unadjusted totals for the civilian noninstitutional population. This reduces but clearly does not eliminate some of the spikes in the growth of *LFCADJL* over history (Figure 3). As seen later, this results in extra variability in the model's prediction for trend growth of the labor force.

[14] The BLS estimates that the introduction of new population controls based on Census 2000 raised the civilian noninstitutional population 16 and older (N16C) and LFC by approximately 2.6 and 1.6 million, respectively. Civilian employment was raised by about 1.6 million at the same time. The aggregate unemployment rate was essentially unaffected by updated population controls based on Census 2000.

[15] The participation rates are usually not affected greatly by the introduction of updated population controls, as the revisions to the totals for the labor force and the civilian noninstitutional population are approximately proportional.

AN ESTIMATE OF TREND GROWTH IN THE LABOR FORCE

Figure 4 displays the growth rate of the civilian labor force after adjustments that smooth the effects of updated population controls, along with a forecast from MA's most recent long-term outlook. The figure also shows the prediction of the trend in the adjusted labor force. The latter includes the version of *LFCADJL* adjusted for revised population controls (the adjustment is admittedly incomplete) and the estimate of "trend" for the behavioral component of the labor force based on the model described previously. Figure 5 shows a corresponding set of estimates for the labor force participation rate.

One of the most obvious features is that the estimate of trend growth is not smooth, especially in history. In part this reflects changes in its behavioral determinants, but it also reflects discontinuities from updated population controls that, given available information, we are able to reduce but not eliminate. The spike in 1990 is an example, as are a pair of sharp declines in the 1960s.

According to the model, trend growth in the labor force peaked in the early 1970s at slightly below 3 percent; but it soon subsided and, for most of the 1980s and the first half of the 1990s, trend growth fluctuated between 1 and 2 percent. It rose briefly in 1996 in response to welfare reform. Declines in the net worth term generated brief increases in the model's prediction for potential labor force growth in the earlier 2000s and again recently (and through the first couple years of the forecast).

Turning to the forecast, trend growth of the labor force is projected to average 0.9 percent from 2008 to 2017, three-tenths of a percentage point higher than in our most recent forecast. The trend in the labor force participation rate is projected to edge down slightly but remain close to 66 percent throughout the forecast through 2017, well above our previous forecast of a decline to 64.6 percent by 2017. The model's predictions are also higher than trend estimates from the Congressional Budget Office (2008) and especially those from Aaronson et al. (2006).

Figure 3

Demographic Contribution to Labor Force Growth with Population Adjustments

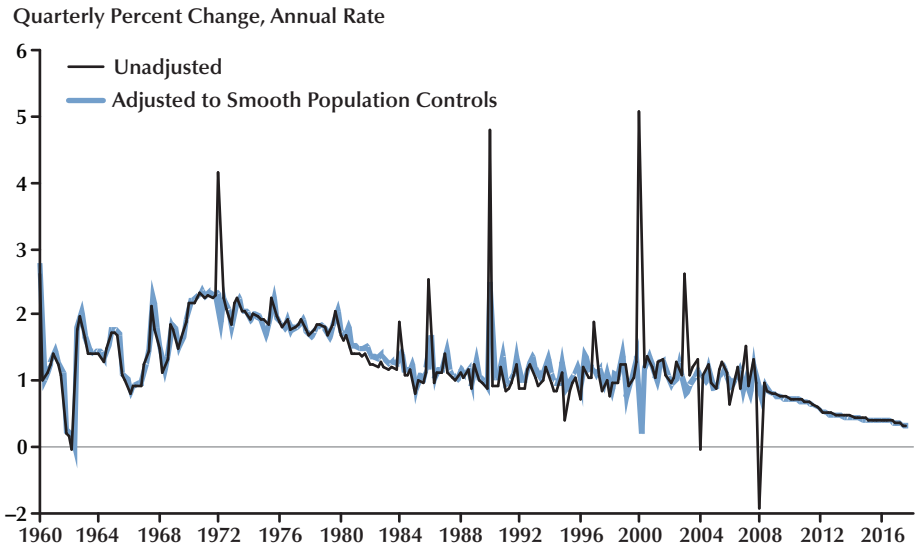


Figure 4

Trend Growth of the Labor Force with Population Control Adjustments

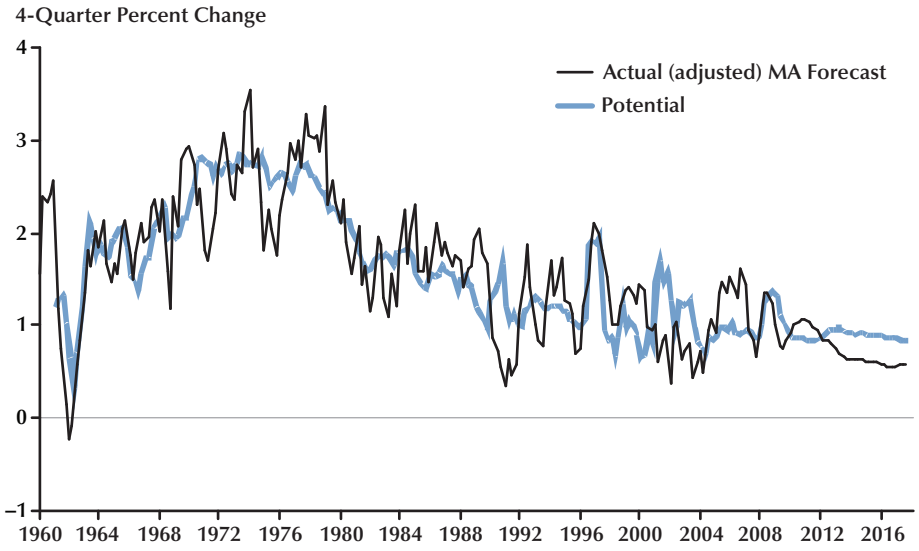
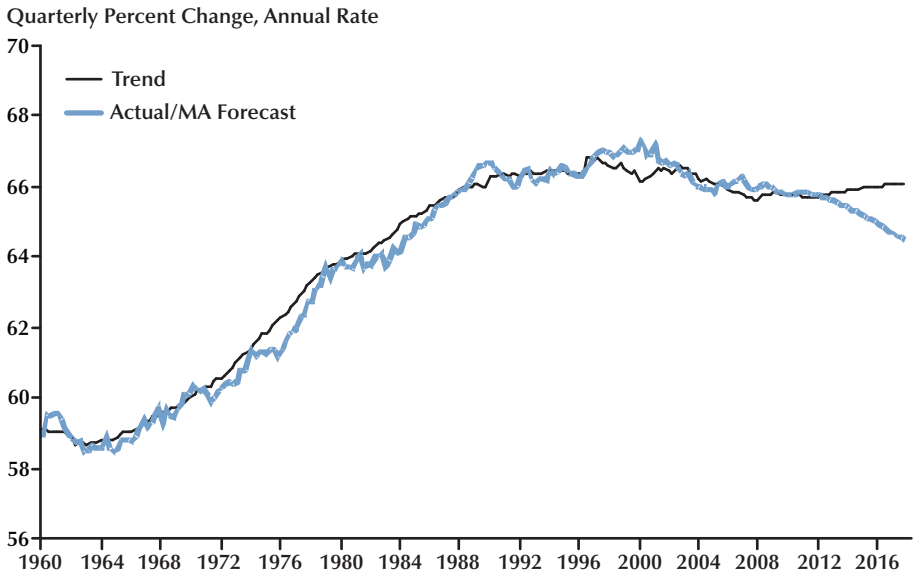


Figure 5

Labor Force Participation (Actual and Trend) with Smoothed Population Controls



MODEL SPECIFICATION DETAILS

We contemplated a larger set of potential behavioral influences on the labor force than those shown in the model in Table 3. Many terms that were considered do not appear in the featured specification because the econometric results did not support their inclusion, including (i) the difference between the marginal and average net-of-tax rates for labor income and the ratio of the marginal to the average net-of-tax rates; (ii) the marriage rate for women; (iii) the ratio of the female to the male participation rate; (iv) the ratio of after-tax Social Security retirement benefits to after-tax hourly labor compensation and the same ratio multiplied by the population share for age 65 and older; (v) a zero-one dummy variable for the elimination in 2000 of the Social Security earnings test for persons who have reached normal retirement age; (vi) replacement of the unemployment rate with separate regressors for the NAIRU and the difference between the unemployment rate and the NAIRU; and (vii) a linear time trend.[16]

Limitations on data availability and labor resources precluded assessing other factors that might influence work/retirement decisions, such as the cost of medical care; parameters that affect Social Security retirement benefits, such as a more nuanced assessment of changes in the earnings test, and changes to the delayed retirement credit; the evolution from defined-benefit to defined-contribution retirement plans; and edu-

[16] Although one of our goals was to develop a behavioral model without relying on ad hoc deterministic trends or shift terms, we did investigate the effect of adding a trend to evaluate whether one or more of the regressors in the featured specification appeared to be significant because it (or they) simply filled the role of a time trend. Fortunately, we did not find that to be the case. When a linear time trend is added to the regression for log(*LFC*/*LFCADJL*), it enters with a negative coefficient and it is borderline statistically significant, with a *t*-statistic of –1.90, while existing level regressors remained statistically significant. The coefficient on the life expectancy term rose by more than one-third and the sum of the contributions from the trend and life expectancy terms in the forecast would have resulted in a prediction for the participation rate that by 2017 is 0.5 percentage points higher than for the featured model. The prediction from the featured model is already stronger than existing forecasts, including our own previous long-term projection, so we are hesitant to adopt specifications that imply even faster labor force growth in the forecast without a *compelling reason to do so*, a hurdle that we did not feel was exceeded with a *t*-statistic of about –1.90 on a deterministic trend term.

cational attainment and involvement. These issues certainly merit further investigation.

The labor force participation rate of young adults in the 16- to 19-year-old age bracket has declined from a peak near 59 percent in the late 1970s to 41 percent as of 2008:Q2. The possibility of further declines in the participation rate of this bracket might constitute a downside risk to projections for the labor force but one that we believe is small. If the participation rate of this age bracket fell during the forecast horizon at a pace comparable to its decline over the past decade (which is steeper than the decline over the entire period from the late 1970s to now), then it would, all else equal, lower the aggregate participation rate in 2017 by approximately 0.4 percentage points. Furthermore, we think the downside risk to the forecast could be even less than suggested by the static calculation. Why? First, our estimation sample, which begins in 1960, includes the entire period of decline in this age bracket, so the model should not be "surprised" by continued declines comparable to those experienced over history. Second, as noted previously, when we added a trend term to the model, the projection for the labor force was actually higher than for the featured model. Third, we tried adding a trend premultiplied by the population share for 16- to 19-year-olds, but it was essentially zero, statistically insignificant (t-statistic of –0.1), and produced no discernible changes in other coefficients or in the model's predictions.[17] Splitting the weighted trend into separate terms for the period up to 1978 and thereafter was equally ineffective. Finally, the decline in the participation rate of 16- to 19-year-olds seems to be related to increasing educational involvement of this group. For 16- and 17-year-olds, school enrollments have risen to more than 95 percent, which presumably leaves relatively little room for additional increases. There might be more room for increased participation for 18- and 19-year-olds, for whom the enrollment percentage has risen to a little over 67 percent.

WHY USE FEMALE LIFE EXPECTANCY DATA?

The rising contribution from the life expectancy term is clearly the most important element of the model that produces a prediction for labor force growth that is higher than in other forecasts. However, we are not inclined to conclude that the model produces an overly optimistic projection for the labor force over the next decade. We are comfortable with the notion that increases in life expectancy raise the amount of wealth required to support a given flow of expenditures in retirement and thereby contribute to increases in the participation rates for older age brackets. Furthermore, other developments are likely to complement the impact of rising life expectancy and contribute to future increases in participation rates in older age brackets, including changes in parameters that influence Social Security retirement benefits, including the ongoing increase in the normal retirement age, the gradual weakening of the earnings test, and the expansion of the delayed retirement credit; rising educational attainment and the increasingly knowledge-based nature of employment; rising costs for health care; the expansion of defined-contribution retirement plans at the expense of defined-benefit plans; and the possibility that employers will adapt to a slowdown in the growth of the population of prime-aged adults by increasing their recruitment and retention efforts for older, skilled workers.

These factors aside, why did we choose the particular form of the life expectancy term—the life expectancy of women at the age of 65, multiplied by the share of women 65 and older in the civilian noninstitutional population ("female-weighted, female life expectancy")? We considered other life expectancy terms, including the male-weighted, male life expectancy at age 65, and the male-weighted, female life expectancy, among others, but we did not include them for several reasons. First, the female-weighted, female life expectancy worked well, with a positive coeffi-

[17] We also considered whether adding a similar term to the model, equal to the population share for the 16- to 24-year-old age bracket times a linear time trend, would change the results. This term did enter with a negative sign when added to the levels regression for log(*LFC*/*LFCADJL*). However, the in-sample predictions were similar to the model without this term, and the out-of-sample forecast projections were virtually identical. Based on this evidence, we chose not to include this term in the model.

cient (as expected) and a high *t*-statistic of nearly 6. Second, *adding* male-weighted life expectancies (for either men or women at age 65) did not materially improve fit and led to similar estimates of the contribution of changes in life expectancy to the growth of the labor force over the forecast period. Third, *replacing* the female-weighted, female life expectancy with either the male-weighted, male life expectancy or the male-weighted, female life expectancy caused the fit to deteriorate somewhat. Fourth, *replacing* the female-weighted, female life expectancy with the male- and female-weighted average life expectancy for men and women at age 65 caused the fit of the equation to deteriorate slightly. Fifth, we found support for a strong role for female life expectancy in a preliminary investigation into the labor force participation rates for specific age brackets of older men *and* women but not for male life expectancy.

Do these statistical results indicate a more important role for female life expectancy is reasonable? We think they do for several reasons. First, except when ill health intervenes, spouses tend to coordinate their work/retirement decisions, suggesting that the decisions of husbands will depend in part on the life expectancy of their wives, and vice versa.[18] Second, on average, women live longer than men, suggesting that the life expectancy of wives is more important for savings and retirement decisions within the household.[19] Third, Goda, Shoven, and Slavov (2007) demonstrate that many individuals experience a sharp increase in their net Social Security tax rate as they age; and, because of the parameters that determine taxes and benefits, on average men are likely to experience a sharper increase than women and at a much earlier age than women. For many men, the sharp increase occurs at or before normal retirement age, creating a financial incentive toward earlier retirement that, on average, is larger for men than for women. This feature of Social Security tends to diminish the role of male life expectancy, and in the context of household decisionmaking, accentuates the importance of female life expectancy for the retirement decisions of both genders.

CONCLUSION: IMPLICATIONS FOR POTENTIAL OUTPUT

The estimate for trend growth of the labor force can be combined with other procedures described in a June 2008 presentation to generate a consistent estimate of potential GDP growth.[20] Here we briefly sketch the implication for potential growth over the forecast through 2017. The main elements of potential GDP are (i) potential growth of hours worked in the nonfarm business sector, (ii) structural productivity growth in the nonfarm business sector, and (iii) trend growth in other GDP. The sum of the first two elements (apart from compounding) provides an estimate of potential GDP growth in the nonfarm business sector. That sector accounts for approximately three-fourths of total GDP.

Trend hours in the nonfarm business sector equals the trend in the workweek, which we assume is roughly flat in the forecast, times potential employment in that sector. The latter is equal to total potential civilian employment less employment in the "other" sectors outside the nonfarm business sector. Our procedures ensure that the "other" employments, which account for roughly 20 percent of total employment, are consistent with our forecasts for the "other" output, which includes government output, output of the household and institutional sectors, and agricultural output. Potential civilian employment is simply one minus the NAIRU (in decimal form) times the potential labor force. Using these methods, we estimate that potential hours growth through 2017 will be close to the estimate of potential labor

[18] Munnell and Sass (2007) discuss many factors that influence the supply of labor for older Americans. They cite several papers showing a strong tendency for husbands and wives to retire within one to two years of each other.

[19] On a related point, consider the work/retirement decisions of widows and widowers. They are likely to be influenced by their own life expectancy but not by the statistical life expectancy of the opposite gender. There are more widows than widowers, which accentuates the role of female life expectancy relative to male life expectancy.

[20] See Matheny's presentation entitled "Research Update: Potential GDP" delivered at MA's June 10, 2008, *Quarterly Outlook Meeting*.

force growth, or about 0.9 percent per annum. This reflects an assumption of a roughly flat trend in the workweek, an essentially constant value for the NAIRU, and an increase in "other" employment that averages about 1.1 percent.

Our estimate of structural productivity growth reflects contributions from capital deepening and growth of total factor productivity. We assume in the forecast that the latter will increase at a 1.2 percent annual rate. Based on projections regarding the growth of capital services in Macroeconomic Advisers' most recent long-term outlook as of this writing, capital deepening is expected to add another 0.8 percentage points to productivity growth in the forecast, resulting in structural productivity growth of about 2.0 percent and potential GDP growth in the nonfarm business sector of about 2.9 percent. Allowing for a contribution from "other" GDP of about 0.4 percentage points on average, this implies that total potential GDP growth would be expected to average about 2.6 percent through 2017. This is two-tenths higher than the Congressional Budget Office's (2008) projection that potential GDP growth will average 2.4 percent over the same period.

REFERENCES

Aaronson, Stephanie; Fallick, Bruce; Figura, Andrew; Pingle, Jonathan and Wascher, William. "The Recent Decline in the Labor Force Participation Rate and Its Implications for Potential Labor Supply." *Brookings Papers on Economic Activity*, Spring 2006, *1*, pp. 69-134.

Bartik, Timothy J. "Displacement and Wage Effects of Welfare Reform," in David Card and Rebecca M. Blank, eds., *Finding Jobs: Work and Welfare Reform*. New York: Russell Sage Foundation, 2000, pp. 72-122.

Blank, Rebecca M. "What Did the 1990s Welfare Reform Accomplish?" Written for the Berkeley Symposium on Poverty and Demographics, the Distribution of Income, and Public Policy, December 2003; updated 2004; http://urbanpolicy.berkeley.edu/pdf/Ch2Blank0404.pdf.

Congressional Budget Office. "The Budget and Economic Outlook: An Update." CBO, September 2008; www.cbo.gov/ftpdocs/97xx/doc9706/09-08-Update.pdf.

Goda, Gopi S.; Shoven, John B. and Slavov, Sita N. "Removing the Disincentives in Social Security for Long Careers." NBER Working Paper No. 13110, National Bureau of Economic Research, May 2007; www.nber.org/papers/w13110.pdf?new_window=1.

Goodstein, Ryan. "The Effect of Wealth on the Labor Force Participation of Older Men." Unpublished manuscript, University of North Carolina–Chapel Hill, March 2008; www.unc.edu/~rmgoodst/wealth.pdf.

Macroeconomic Advisers. *Long-Term Economic Outlook*. September 24, 2008.

Munnell, Alicia H. and Sass, Steven A. "The Labor Supply of Older Americans." Working Paper No. 2007-12, Center for Retirement Research at Boston College, June 2007; http://crr.bc.edu/images/stories/Working_Papers/wp_2007-12.pdf?phpMyAdmin=43ac483c4de9t51d9eb41.

Matheny, Ken. "Research Update: Potential GDP," presented at Macroeconomics Advisers' Quarterly Outlook Meeting, June 10, 2008, St. Louis, Missouri.

Social Security Administration. *Personal Responsibility and Work Opportunity Reconciliation Act of 1996* (Public Law 104-193; §115.42 U.S.C. 862a), August 22, 1996; www.ssa.gov/OP_Home/comp2/F104-193.html.

U.S. Census Bureau. *Current Population Survey*. www.census.gov/cps/.

Commentary

Ellis W. Tallman

Macroeconomists ranging from policymakers to business and economic forecasters use the concept of potential output in specific economic constructs. In some applications, economists look at the "output gap"—the difference between an estimate of potential output and the measure of actual real output—as a forecasting tool for inflation to gauge whether deviations of real output from potential should lead to increases or decreases in future inflation. Monetary policymakers use potential output in this way in applications of the Taylor rule framework. Separately, economic forecasters use the estimate of potential output as a comprehensive measure of the underlying trend in real output growth for the economy. In the latter usage, calculating an estimate for potential output typically starts with estimates of the primary factors of production—capital and labor inputs.

The motivation for the paper "Trends in the Aggregate Labor Force" (Matheny, 2009) is the search for a more accurate and comprehensive measure of the labor input for potential output estimates. The goal is commendable, and there are few reasons to fault the author for committing resources toward producing an improved estimate for the labor input. Matheny uses a more detailed set of labor data series from which to calculate an estimate of the available labor force and ultimately to create an estimate of the labor input measure. Even in a preliminary form, the paper provides a concise survey of a work in progress as it outlines a number of additional issues that remain unsettled. Among the main findings is an influential role of factors that could influence the labor force participation of women 55 and older as inferred from an estimated regression model. This participation rate has increased over time and is currently higher than has been observed historically. The bottom line from the research is that estimates of potential output that do not take into account behavioral responses that reflect increasing labor force participation rates of the older population will *under*estimate the growth in the labor force and thereby underestimate the growth rate for potential output.

In this discussion, I focus my comments on these central findings of the research. First, my discussion outlines the contribution of the paper with respect to the calculation of the demographic component of the labor force. Next, the comments focus on the main explanatory variable in the aggregate labor force participation rate regression—the population proportion of women 65 and older weighted by life expectancy of women at age 65 (the behavioral variable *WT65F_LEF65* in the paper). Next, the discussion investigates whether other, additional factors may explain the strong observed correlation between the dependent variable (a change in the aggregate labor force participation rate) and the *WT65F_LEF65* variable. More narrowly, I ask whether there are underlying variables that may explain the increased labor force participation of women 65 and older in addition to the rising life expectancy of women.

Ellis W. Tallman is the Danforth-Lewis Professor of Economics at Oberlin College.

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 311-16.
© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

Further, the discussion investigates whether the implied elasticity of labor force participation with respect to the *WT65F_LEF65* variable in the regression is consistent with feasible changes in the labor force participation rate of women 65 and older. The findings suggest that there remain numerous interesting research questions that these observations raise for labor economists in particular. Finally, I make some suggestions for broadening the appeal of the work.

CALCULATION OF THE LABOR INPUT

The bottom-line finding of the paper is that a revised labor input measure contributes an increase of nearly 0.5 percentage points to the estimation of potential output growth. The measure sounds small, but that kind of calculation is significant, especially if it is an accurate forecast. Clearly, the labor input for the estimation of potential output is only one of several inputs important for that calculation. Rather than highlighting the limitation of focusing on only one factor input, this discussion adopts the view, as stated in the paper, that refining the labor input measure for a potential output estimate is "low-hanging fruit."

The treatment of labor force growth is central to the paper, and it clarifies the distinction between the components of labor force growth that reflect only shifting population demographics and those that reflect labor force participation rates of the demographic subcategories (gender and age categories). The population demographics can be predicted reliably from population data. In contrast, the labor force participation rates may vary as a result of changes in economic situation, life expectancy, and so on and therefore may deviate from a trend labor force participation rate. The paper makes a notable contribution to the measurement of the labor input estimate from the calculation of additional gender/age brackets and the incorporation of the related labor force participation rates. Specifically, the paper increases the number of age brackets from 7 to 15, thereby increasing the detail of the population characteristics and likely affording a more comprehensive labor force estimate. Further, the paper uses the narrower population measure—civilian noninstitutional population—rather than resident population data—to generate more precise estimates of labor force. Using available population demographic data (civilian noninstitutional population), the author calculates a chain index of the age-and-gender population detail at the quarterly frequency.

The labor force series uses the participation rates from the previous period (t–1) as weights for the population demographics for each age and gender category in the current period and thereby emphasizes the impact of demographic factors. The series, listed as *LFCADJL*, measures the quarter-to-quarter growth as entirely due to demographic factors. The previous description understates the amount of meticulous data analysis required to formulate an improved labor force growth estimate.

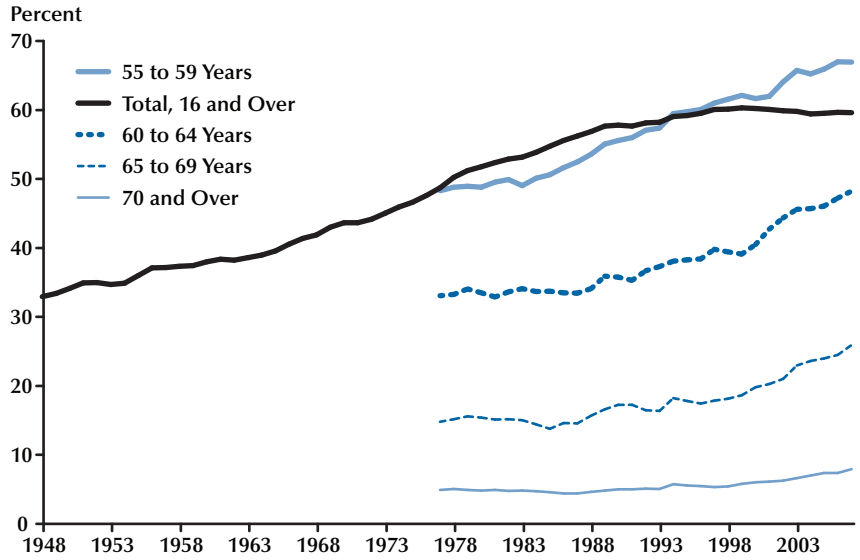
The influence of population growth in a given demographic component on the labor force relies on the proportion of that demographic group in the labor force (noting the dating differences of the aggregate and age-gender bracket). Clearly, if a demographic group—like those 75 and older—grows rapidly, but the share of that demographic group in the labor force is low, then the influence of that population growth on the labor force is small. As noted previously, this labor force measure highlights the demographic components of population and its effects on the labor force if labor force participation rates were not changing.

The accounting aspect of the investigation, that is, the addition of demographic subcategories in the labor input measure, provides only the groundwork for the economic analysis of the behavioral element of the labor force input. Still, the general work on the comprehensive dataset offered opportunities to investigate the labor force participation rates of various age and gender brackets.

Figure 1 illustrates the specific isolation of the labor force participation rate for women 55 and older and its subcategories (55-59, 60-64, 65-69, and 70 and older). The observation of rising labor force participation rates of women 65 and older compels further investigation, and the empirical work investigates whether including a

Figure 1

Female Labor Force Participation by Age



measure of the life expectancy of women at age 65 multiplied by the population proportion of women 65 and older adds explanatory power to a regression to forecast the behavioral element (labor force participation rates) of the aggregate labor input. The research provides an interesting initial inquiry into a regression-based empirical model to explain (and then predict) the aggregate labor force participation rate.

THE REGRESSION

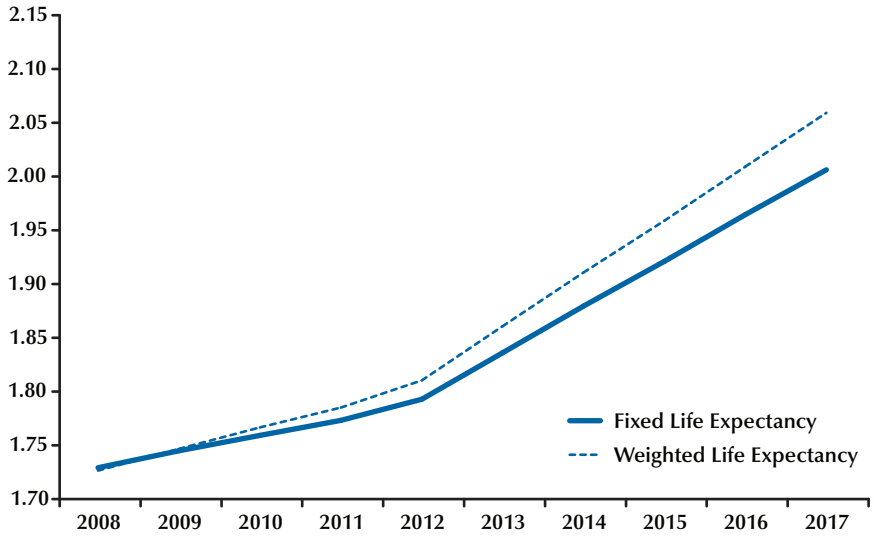
The regression analysis in the paper uses a set of explanatory variables intended to account for the behavioral changes in the aggregate labor force participation rate. The paper outlines and describes the regression in detail; my discussion here focuses on one key explanatory variable—life expectancy of women at age 65 times the share of women age 65 and older in the adult population. This variable is especially important for the forecast period 2011-17 and largely explains the increase in labor force participation in the new estimate for the labor input.

The finding raises a number of questions; the main one is whether a regression model that is meant to explain the behavioral variations in aggregate labor force participation rates attributes too much influence to this particular variable. It would be helpful to have an explicit accounting for the quantitative increase in the labor force generated by increases in *WT65F_LEF65*. First, the explanatory series should have a positive effect on the participation rates of women 65 and older. Second, the increase in the participation rate of women 65 and older times the population of women 65 and older should generate an increase in the labor force of women 65 and older of a similar magnitude to the one generated by the aggregate labor force participation rate regression.[1] Conversely, the author can work in the opposite direction by taking the increase in the labor force implied by the aggregate labor force participation rate regression coefficient and investigate the required increase in the labor force participation

[1] A related question is whether the rise in life expectancy for women at age 65 has significant explanatory power for the participation rate of women 65 and older.

Figure 2

Comparison of Weighted versus Unweighted Population Proportion of Women 65 and Older



SOURCE: Population projections from www.bls.gov/emp/emplab1.htm.

rate for women 65 and older that would be necessary to generate the labor force observation.

Second, the variable itself is composed of two increasing components—the population proportion of women 65 and older and the life expectancy of women at age 65. Figure 2 shows the estimated series for 2008-17 along with a series in which the life expectancy after age 65 is held fixed at 19.7 years (the expectancy in 2008). Clearly, the dominant component of the series is the population proportion of women 65 and older, which reflects the demographic influence of the large baby boom generation. If the life expectancy component of the measure were important to the regression results, then a regression using only the population proportion of women 65 and older should not have much explanatory power. If, on the other hand, the regression results are similar, then the result suggests that behavioral variations in the aggregate labor force participation rate respond to demographic movements. Such an explanation would be unsatisfying.

The author could also try a few other techniques to assess the feasibility of the result. The data on the population of women 65 and older could be used to carry the demographic analysis out to the forecast year 2017, given standard assumptions for the mortality rate, and so on. Then, the analysis can focus on examining a set of possible labor force participation rates for women 65 and older and how different labor force participation rates affect the aggregate labor force. For example, a particular labor force participation rate for this specific entry could be chosen to determine what that participation rate suggests for the aggregate labor force calculations. The accounting of the population demographics is noncontroversial; the examination of the labor force implications of various labor force participation rates for this demographic can be thought of as a conditional forecasting exercise. The analysis would allow an inference for (i) whether the explanatory power of the life expectancy of women at age 65 reflects only the contribution of women 65 and older to the labor force or (ii) whether the measure reflects further influences as a proxy. Other factors may be correlated with the specific regressor variable; that result, if found,

would allow further refinement of the initial finding. Additional research would then aim at uncovering the additional factors with the goal of identifying (or at least clarifying) other underlying sources for the increase in the labor force participation rate.

The regression model is meant to explain the behavioral aspects of labor force participation, although the current findings also introduce some intriguing questions that, the author admits, remain unsettled. Some of these questions are addressed in the paper. For example, the author investigates whether aggregate wealth calculations explain the increased labor force participation; initial results suggest that a measure of wealth was not associated with the increase in labor force participation. The result may be only preliminary, however, because it uses an aggregate measure of per capita wealth. In accord with the previous suggestions, an analysis of disaggregate wealth measures that relate to specific demographic groups—for example, the population 65 and older—may have explanatory power for the labor force participation of that subcategory.

Increased life expectancy of women at age 65 may explain the higher-than-anticipated labor force participation of women 65 and older; it makes intuitive sense. Separately, there may be important cost-of-living elements that drive a higher labor force participation rate for those 65 and older. Recent empirical work by Broda and Romalis (2008) suggests that economic analysis can be more precise with respect to the "wage gap" with more precise price deflators that relate more closely to the prices and to the expenditure patterns of the relevant income groups in the comparison. Perhaps a similar approach can be used for the population 65 and older. The consumer basket for a person 65 or older could be notably different from the standard basket of goods used in the calculation of the consumer price index. One might expect a larger component of spending on prescription drugs and for health services for those 65 and older; then, there might be a faster rate of inflation for that cohort than for the general public. A rising cost of living for those facing fixed incomes might lead to a higher-than-expected rate of labor force participation. In this case, longer work lives may also be related to the increased life expectancy of those 65 and older.

These comments and criticisms aim to refine and dissect a notable result. The basic finding of the regression highlights a major flaw in the use of fixed or trend participation rates in the calculation of "potential" labor force. That contribution remains even though several other factors remain to be investigated as potential sources for a forecast of increased labor force participation in the aggregate labor input measure. Specifically, the empirical work captures some of the observed changes in the labor force decisions of older individuals and the effect of those changes on the labor force projections for the future. The point is especially important given the demographic impact of the baby boom generation on the labor force as that generation approaches retirement age. If the baby boomers stay in the labor force longer than anticipated, there will be important labor market effects, and this paper emphasizes that point.

IDEAS FOR ILLUSTRATING THE IMPORTANCE OF THE LABOR INPUT REVISION

The paper provides ample evidence to suggest that the labor force participation rate increase among those aged 65 and older may increase the potential labor force above the pessimistic forecasts offered by the demographic data alone. Yet, the labor input is only one component of the calculation of potential output. In addition, some influential treatments of estimating potential output have instead focused on the calculation of the effect of computers on economic growth (Jorgensen, 2005). The paper can use the impending baby boomer event to motivate the relevance of the labor input in the calculation of a real-time potential output estimate. The estimate of potential output that incorporates revised labor force participation rates (new behavioral labor force estimates) displays a deviation from the previous potential gross domestic product estimate that is

larger than at any earlier period in the estimation sample.

Revisions to potential gross domestic product measures have been the subject of numerous empirical investigations (Orphanides, 2001); the paper could incorporate some of these findings to illustrate where prior estimates of potential output failed to account for certain factors. It is likely that the current labor force participation rates are undergoing an adjustment that, in retrospect, will seem more apparent.

It may be worthwhile, though not necessarily for this research agenda, to determine whether there are precedents for the labor force participation rate underestimate. Perhaps the increase in female labor force participation through the 1970s and 1980s was relatively unexpected. More recently, the influence of immigration may have affected estimates of the labor input. The paper can highlight further its relevance if it can isolate historical episodes in which more accurate labor input measures for a potential output estimate were empirically important.

CONCLUSION

The paper offers an interesting contribution to the calculation of the labor input for a potential output estimate by increasing the disaggregation of the demographic components of the labor force input. Further, the paper provides initial results for a model of the behavioral element of the labor force input, essentially, a model of the aggregate labor force participation rate. The data-based enhancements for the labor input measure are noncontroversial and should offer a roadmap for other estimates of potential output growth. The model-based predictions regarding the aggregate labor force participation rates are intended to stimulate discussion rather than be taken as ultimate findings. The discussion highlights a number of avenues to pursue to refine our understanding of the estimated regression model and to assess its robustness.

The overall implication of the regression analysis suggests that the pessimistic forecasts of labor force growth in the United States may be too low, and that suggestion contributes to an interesting debate about labor force dynamics in the medium term. The paper raises a number of interesting research topics from the aggregate labor data. Perhaps other interesting research could use the aggregate research results as motivation for modeling the behavioral decisions for labor force participation on the level of the disaggregate population demographics. Although these ideas are not part of the author's research agenda, labor economists could offer findings that then help isolate additional sources of the increased labor force participation rate.

REFERENCES

Broda, Christian and Romalis, John. "Inequality and Prices: Does China Benefit the Poor in America?" Unpublished manuscript, University of Chicago, May 2008; http://faculty.chicagogsb.edu/christian.broda/website/research/unrestricted/BrodaRomalis_TradeInequality.pdf.

Jorgenson, Dale W. "Accounting for Growth in the Information Age," in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*. Volume 1A. Chapter 10. Amsterdam: Elsevier, 2005, pp. 743-815; www.economics.harvard.edu/faculty/jorgenson/files/acounting_for_growth_050121.pdf.

Matheny, Kenneth J. "Trends in the Aggregate Labor Force." Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 297-309.

Orphanides, Athanasios. "Monetary Policy Rules Based on Real-Time Data." *American Economic Review*, September 2001, *91*(4), pp. 964-85.

Potential Output in a Rapidly Developing Economy: The Case of China and a Comparison with the United States and the European Union

Jinghai Zheng, Angang Hu, and Arne Bigsten

The authors use a growth accounting framework to examine growth of the rapidly developing Chinese economy. Their findings support the view that, although feasible in the intermediate term, China's recent pattern of extensive growth is not sustainable in the long run. The authors believe that China will be able to sustain a growth rate of 8 to 9 percent for an extended period if it moves from extensive to intensive growth. They next compare potential growth in China with historical developments in the United States and the European Union. They discuss the differences in production structure and level of development across the three economies that may explain the countries' varied intermediate-term growth prospects. Finally, the authors provide an analysis of "green" gross domestic product and the role of natural resources in China's growth. (JEL L10, L96, O30)

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 317-42.

The rapid development of emerging markets is changing the landscape of the world economy and may have profound implications for international relations. China has often been regarded as the most influential emerging market economy. Projections indicate that the absolute size of the Chinese economy may be larger than that of the United States within two to three decades. While China's growth performance since reform has been hailed as an economic miracle (Lin, Cai, and Li, 1996), concerns over the sustainability of its growth pattern have emerged in recent years when measured total factor productivity (TFP) growth has slowed.

In recent years, economists have increasingly referred to China's growth pattern as "extensive." Extensive growth is intrinsically unsustainable because growth is generated mainly through an increase in the quantity of inputs rather than increased productivity. In a previous paper (Zheng, Bigsten, and Hu, 2009), we focused on China's capital deepening versus TFP growth and private versus government initiatives. In this article, we first compare China's growth performance with what would otherwise have been feasible, taking into account the main factors commonly employed to generate growth in rapidly developing economies. In other words, we compare official statistics with estimates of "potential" output growth to shed further light on China's recent growth patterns.

Jinghai Zheng is a senior research fellow in the Department of International Economics at the Norwegian Institute of International Affairs, Norway, an associate professor in the department of economics at Gothenburg University, Sweden, and a guest research fellow at the Centre for China Studies, Tsinghua University, Beijing, China. Angang Hu is a professor at the School of Public Policy and Management at Tsinghua University. Arne Bigsten is a professor in the department of economics at Gothenburg University. The authors thank Justin Yifu Lin for his support and encouragement of this project and Xiaodong Zhu for useful discussion. The study was also presented at the Chinese Economic Association (Europe) Inaugural Ceremony, December 17, 2008, Oslo, Norway. The authors thank participants at the event and especially those who commented on their paper. The study benefited from research funding from the Center of Industrial Development and Environmental Governance at the School of Public Policy and Management, Tsinghua University. Yuning Gao provided excellent research assistance.

© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

Second, we provide projections of the future potential of the Chinese economy and discuss China's impact on the world economy. Specifically, we compare potential growth in China with that for the United States and European Union (EU). We note that structural characteristics, rapid accumulation in capital stock, and improvement in labor quality are the major factors behind China's phenomenal economic growth. China's future TFP growth is likely to be faster than that of the United States and EU because of the stock of world knowledge it may easily access at affordable prices to enhance its production possibilities (Prescott, 2002).

Nobel laureate Ed Prescott (1998) asked why "growth miracles" are a recent phenomenon. We suspect that the main reasons are differences in production structure and in the level of development. Examples include the East Asian newly industrialized countries (NICs), to some extent post-WWII Japan and Germany, and the Soviet Union between the first and second World Wars and in the early years of the Cold War. Now, due to rapid industrialization, China will soon join the ranks of the high-performing East Asian nations. Understanding the causes and conditions of economic miracles may prove useful for developing countries. Understanding differences in production structure and the level of development may also help explain why productivity slowed in the United States and EU in the early 1970s, then started to surge in the United States but stagnated in Europe in the mid-1990s.

To analyze growth potential, we consider the usual suspects of demographics, rural-urban migration, and aging. In addition, we discuss how estimates of potential output have affected Chinese government policy regarding growth planning. Because environmental regulations and concerns are of increasing international importance, we assess in the final section of this analysis the influence of environmental factors—specifically, to what extent past economic growth reflected environmental "inputs" not elsewhere accounted for.

THE ANALYTICAL FRAMEWORK

Years before the current worldwide credit crunch, the economics literature included many works that foresaw the looming economic crisis (e.g., Gordon, 2005; Phelps, 2004; Stiglitz, 2002; and Brenner, 2000 and 2004). Gordon's (2005) application of the growth accounting framework to the study of the U.S. productivity revival and slowdown stands out as convincing evidence that economic theory can powerfully inform empirical analysis for macroeconomic planning.

Since the publication of Solow's seminal work on technical progress and the aggregate production function, growth accounting has been used to assess the economic performance of the former Soviet Union (Ofer, 1987), raise concerns about the sustainability of the economies of the East Asian "tigers" (Hong Kong, South Korea, Singapore, and Taiwan) just a few years before the East Asian financial crisis (Young, 1995; Kim and Lau, 1994; and Krugman, 1994), and, recently, forewarn planners about the macroeconomic imbalances in China (Zheng and Hu, 2006).

Adequately implemented and understood, growth accounting is a useful instrument for improving the analysis of growth potential for many countries and regions. Several examples in the literature show that growth accounting methods are sensitive enough to detect significant changes in productivity performance if production parameters are carefully chosen.

Growth accounting decomposes growth in output into its components:

$$\frac{\dot{Y}}{Y} = \frac{\dot{A}}{A} + \alpha\frac{\dot{K}}{K} + (1-\alpha)\frac{\dot{L}}{L}, \tag{1}$$

where Y is gross domestic product (GDP) and $\dot{Y}$ change in GDP over time; K is capital stock and $\dot{K}$ the change in capital stock; labor is L and $\dot{L}$ the change in labor input; TFP growth is $\dot{A}/A$; $0 < \alpha < 1$ is the output elasticity of capital; and $(1 - \alpha)$ is the output elasticity of labor.

Potential output growth may be calculated via equation (1) from knowledge of the potential growth of each of the right-hand side components, plus estimates of output elasticities for the various inputs. Obviously, both the growth potentials

and the output elasticities will differ among countries, reflecting structural differences. Typical growth accounting structures are represented as follows: For China (Chow and Li, 2002; and Chow, 2008),

$$\frac{\dot{Y}}{Y} = \frac{\dot{A}}{A} + 0.6\frac{\dot{K}}{K} + 0.4\frac{\dot{L}}{L}; \tag{2}$$

for the United States (Congressional Budget Office [CBO], 2001),

$$\frac{\dot{Y}}{Y} = \frac{\dot{A}}{A} + 0.3\frac{\dot{K}}{K} + 0.7\frac{\dot{L}}{L}; \tag{3}$$

and for the EU (Mossu and Westermann, 2005),[1]

$$\frac{\dot{Y}}{Y} = \frac{\dot{A}}{A} + 0.4\frac{\dot{K}}{K} + 0.6\frac{\dot{L}}{L}. \tag{4}$$

China has an output elasticity of capital of 0.6, compared with 0.3 for the United States. Differences of this magnitude are large enough to generate a significant difference between the growth potential of the two economies. For example, a capital stock growth rate of 10 percent would enable China to grow by at least 6 percent per year, whereas, all else constant, it would increase the U.S. growth rate by only 3 percent per year. Growth differences can also be related to differences in investment in physical capital as well as in TFP growth. For developing economies such as China, investment opportunities abound because of the country's relatively low level of development compared with that of the United States, EU, and other industrialized countries. For the same reason, China more easily absorbs and benefits from existing worldwide technology, whereas developed economies, especially the United States, have to rely on new knowledge and innovations to shift their production frontier.

Steady-State and Sustainable Growth

The growth accounting framework provides a compact formula for the study of potential output growth. We define "potential output" as the highest level of real GDP that can be sustained over the period of interest. Growth associated with potential output can therefore be termed "sustainable growth." We divide sustainable growth into three categories according to the different time frames considered.

The first concept of sustainable growth refers to circumstances in which certain measures of output growth are maintained permanently as time goes to infinity. The literature offers two different though related output measures that may be used in this context. Different studies have used them for different purposes. Sustainable growth can be defined as a growth pattern that generates sustained growth in per capita income over an infinite time horizon. Usually, per capita income is treated as a measure of living standards. Following Romer (2006), the standard Solow growth model can be expressed as follows:

$$\begin{aligned} &Y = F\big(K, L, A(t)\big), \quad \dot{A}(t) \geq 0, \\ &A(t) = e^{xt}, \quad L(t) = e^{nt}, \end{aligned} \tag{5}$$

where Y is total output, $F(\cdot)$ is the production function, K is capital input, L is labor input, and $A(t)$ is the level of technology that progresses at the exponential rate x while the labor force grows at the exponential rate n. The change in capital stock is given by

$$\dot{K} = I - \delta K = s \cdot F\big(K, L, A(t)\big) - \delta K, \tag{6}$$

where I is investment, δ the real depreciation rate, and s the saving rate. For $I = sY = s \cdot F(K,L,A(t))$, we have

$$\dot{K} = s \cdot F\big(K, L, A(t)\big) - \delta \cdot K. \tag{7}$$

Dividing by labor input, L, on both sides of equation (7) yields

$$\frac{1}{L}\frac{dK}{dt} = s \cdot F\left(\frac{K}{L}, A(t)\right) - \delta\frac{K}{L}. \tag{8}$$

Because $k = K/L$, the growth rate of k can be written as

$$\dot{k} = \frac{d(K/L)}{dt} = \frac{1}{L}\frac{dK}{dt} - \frac{K}{L^2}\frac{dL}{dt} = \frac{1}{L}\frac{dK}{dt} - \frac{K}{L}n, \tag{9}$$

where $n = \dfrac{\Delta L}{L}$.

[1] Proietti, Musso, and Westermann (2007) set capital elasticity at 0.35 and labor elasticity at 0.65.

Rearrange equation (9):

$$\frac{1}{L}\frac{dK}{dt} = \dot{k} + nk = s \cdot F\big(k, A(t)\big) - \delta k; \tag{10}$$

combine equations (8) and (10):

$$\dot{k} = s \cdot F\big(k, A(t)\big) - (n+\delta) \cdot k; \tag{11}$$

then divide k on both sides of the equation to get the growth rate of k given by

$$\gamma_k = s \cdot F\big(1, A(t)/k\big) - (n+\delta). \tag{12}$$

At steady state, γ_k^* is constant, which requires that s, n, and δ are also constant. Thus the average product of capital, $F(k,A(t))/k$, is constant in the steady state. Because of constant returns to scale, $F(1,A(t)/k)$ is therefore constant only if k and $A(t)$ grow at the same rate; that is $\gamma_k^* = x$. Output per capita is given by

$$y = F\big(k, A(t)\big) = k \cdot F\big(1, A(t)/k\big), \tag{13}$$

and the steady-state growth rate of $y = x$. This implies that, in the long run, output per capita grows at the rate of technical progress, x. Note that this conclusion is conditioned on parameters of the model staying constant, including the saving rate and, hence, the rate of capital formation. This property of the model may explain why developing economies can grow faster, as exhibited by the growth miracles in the East Asian NICs, than developed economies because the potential for absorbing new technologies is larger in the former.

Another important implication of the Solow growth model is that less-advanced economies, such as China, will tend to have higher growth rates in per capita income than the more-advanced economies, such as the United States and EU, because there are more investment opportunities in developing nations. The World Bank (1997, p. 12) called this phenomenon "the advantages of the backwardness." This property is also referred to as "absolute convergence" when the analysis is not conditioned on other characteristics of the economies and "conditional convergence" when the analysis is only valid among economies with the same steady-state positions. However, caution should be exercised when applying this property of the model to real-world situations. The property demonstrates only what the supply side of the economy could achieve if other factors, such as demand conditions, efficiency of the economy, and political stability, are present.

Extensive versus Intensive Growth

Sustainable growth might as well be interpreted as growth of GDP only, which is particularly interesting if one is interested in the absolute size of the economy. It is the size of the aggregate economy that matters with regard to international influence in both economics and politics. Sustainable growth in this context means the rate of investment need not rise in order to maintain a given rate of GDP growth. Such sustainable growth is considered *intensive* growth. A borderline case for sustainable growth of this kind is when the capital stock growth rate equals the GDP growth rate.

Extensive growth refers to a growth strategy focused on increasing the quantities of inputs (Irmen, 2005). As capital accumulation and growth of the labor force raise the growth rate of aggregate output, because of diminishing returns these growth effects will not have a permanent effect on per capita income growth (Irmen, 2005). In contrast, intensive growth focuses on TFP. In our model, labor growth and TFP growth are exogenous; the only input with endogenous growth is capital. A key feature of the extensive growth model is that capital grows faster than GDP (or gross national product) because of the high growth rate of capital on the one hand and few productivity advancements on the other. Consequently, the share of investment in GDP, in constant prices, must grow continuously to sustain the growth rate of capital (Ofer, 1987). Specifically, the relation between I (investment), K (the capital stock), and Y (national product) in real terms can be written as follows:

$$I/K = (I/Y)(Y/K). \tag{14}$$

Notice that the growth accounting formula is given by $\dot{Y}/Y = \dot{A}/A + \alpha\dot{K}/K + \beta\dot{L}/L$, where $(\cdot)$ denotes growth rates, L is labor, and A is the level of technology. Given the growth rate in the labor input and the rate of technological progress, sustainable growth in Y requires sustainable growth

in K. Under intensive growth, $\dot{K}/K < \dot{Y}/Y$, so Y/K rises over time. For $I/K(= \dot{K}/K)$ to stay constant, I/Y must decline; that is, $\dot{I}/I < \dot{Y}/Y$. In other words, the gross capital formation rate does not have to rise to sustain a given growth rate in output, which is feasible.

Under extensive growth, $\dot{K}/K > \dot{Y}/Y$, so Y/K declines and a constant I/K implies a rising I/Y, which is not sustainable in the long run. Moreover, the share of investment in GNP in current prices may be written as $I_C/K_C = IP_I/YP_Y$, where C represents "in current prices" and P the price level. A change in the relative price of I (for example, due to faster technological change) may slow the rise of I/Y in real terms.

If the initial capital stock growth rate is sufficiently low, the economy will grow for a sustained period of time even if capital stock growth exceeds GDP growth substantially. Examples are the Soviet economy in the 1950s and 1960s, the Japanese economy during about the same period, and later the East Asian NICs from the 1960s to 1980s. These economies all experienced rapid economic growth in a relatively short period of time. If the capital growth rate is too high, extensive growth may not be sustainable in the intermediate term or the long run. In some typical cases, sustainable growth requires that the saving rate (and hence the capital formation rate) vary only within a feasible range (say, between 0 to 50 percent) if borrowing is not allowed. Compared with sustainable growth in per capita income in the long run, this type of growth can be sustained only for a limited time because it relies on growth of transitional dynamics rather than on steady-state growth capabilities.

Sustainable Growth with Input Constraints

A third concept of sustainable/potential growth is related to sustainable growth in inputs, especially labor. Everything else equal, change in the labor input can be crucial for growth to be sustained. The economic history of many countries shows that demographics are important for rapid economic growth. In many developed countries, faster growth rates may not be sustained simply because labor is lacking. A country with a large population either too young or too old to work will have a lower growth rate than one with a large working-age population.

Following Musso and Westermann (2005), we decompose labor input into its components. Because we do not have hours worked as a measure of labor, we use employment. Employment, E, at time t is defined as the difference between the labor force, N, and total unemployment, U, and can be expressed as a function of the unemployment rate, ur. The labor force is the product of the participation rate, pr, and the working-age population, p^{WA}. The working-age population is a function of total population, P, and the dependency ratio, dr, where the latter is defined as the ratio between the number of persons below 15 and above 64 years of age and the working-age population. These relationships are summarized as follows:

$$\begin{aligned} E_t &\equiv N_t - U_t = N_t \cdot (1 - ur_t) \\ N_t &\equiv pr_t \cdot P_t^{WA} \\ P_t^{WA} &\equiv P_t \cdot \frac{1}{1 + dr_t}. \end{aligned} \tag{15}$$

The potential GDP growth of China may be expressed as

$$\begin{aligned} g_Y &= g_A + \alpha(i - \delta) + (1 - \alpha) \\ &\left(g_h - \frac{ur}{1 - ur} g_{ur} + g_{pr} - \frac{dr}{1 + dr} g_{dr} + g_p \right), \end{aligned} \tag{16}$$

where the variables are

g_Y, the growth rate of potential output, Y;
g_A, growth of total factor productivity, A;
α, the output elasticity of capital;
$(1 - \alpha)$, the output elasticity of labor;
i, the investment rate;
δ, the depreciation rate;
g_h, growth in years of schooling, h;
g_{ur}, growth of the unemployment rate, ur;
g_{pr}, growth of the participation rate, pr;
g_{dr}, growth of the dependency ratio, dr; and,
g_p, the growth rate of the population, p.

A Comparison of the Three Concepts

A comparison of the three concepts above can be derived from the Solow growth model, but each emphasizes a different aspect of the growth problem. Sustainable growth is derived according to the steady-state solution of the dynamic system. It refers to the mathematical long run, given that the saving rate, depreciation, and population growth are fixed. The variable of interest is the growth rate of per capita income and, hence, TFP growth. In this case, the Solow growth model predicts that low-income economies will grow faster than high-income economies, which leads to the concept of convergence.

The second concept, extensive versus intensive growth, concerns directly the GDP growth rate rather than per capita income. In this case, the saving rate is allowed to change. When the investment rate is so high that the saving rate must rise to, say, over 50 percent of GDP, then the growth pattern is considered problematic. Such growth is not sustainable, even in the intermediate term. However, the problem may arise slowly: If the capital stock growth rate is only 3 percent per year, it may take two or three decades for the saving rate to rise to 30 percent of GDP, even if the growth pattern was initially extensive. This is a major difference between rapidly developing and developed economies. The latter need not worry much about the intermediate term if growth in capital stock exceeds GDP growth, because growth rates are generally low in the relevant variables. In the long run, however, no extensive growth pattern is sustainable. This concept heightens the need to pay attention to the pattern of capital accumulation.

The third concept emphasizes the input constraints. Growth will be sustained only as long as sufficient inputs are available at a given point in time. This formulation is often used to separate the labor input into its components.

EMPIRICAL RESULTS

In this section, we present two sets of empirical results within the framework outlined previously. We first use data from 1978-2007 to update our growth accounting results from Zheng, Bigsten, and Hu (2009). The emphasis is the time-series behavior of TFP growth. Based on our TFP estimates, we provide potential growth measures conditional on the given investment rate and factors related to labor input, including demographics and the labor participation rate. We then compare the estimated potential growth with official statistics.

The second set of results offers projections of future growth. The growth scenarios should not be seen as simple extrapolations based on historical data. In fact, our calibration exercise relies heavily on knowledge of the production structure of the Chinese economy and the concept of intensive growth.

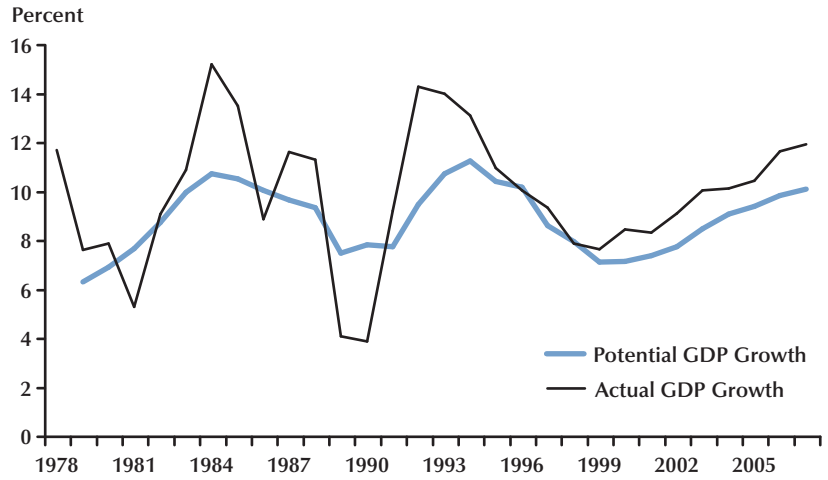
China's Growth Pattern and Potential

China's development strategy in recent years has been successful in promoting rapid economic growth, but it also created a series of macroeconomic imbalances. The rapid growth has benefited China both economically and politically, but whether it can or will be sustained, and for how long, is uncertain. The growth has been generated mainly through the expansion of investment (extensive growth) and only marginally through increased productivity (intensive growth). Some economists fear that if corrective measures are not taken, per capita income will eventually cease to grow. Kuijs and Wang (2006) point out that, if China's current growth strategies are unchanged, the investment-to-GDP ratio would need to reach unprecedented levels in the next two decades in order to maintain GDP growth of 8 percent per year. Our estimates in Table 1 show that China's growth pattern has been extremely extensive, with capital stock growth exceeding GDP growth by 3.56 percent during 1995-2007.

Next, we use equation (16) to calculate a measure of potential growth during 1978-2007. Our measure is built from estimates of the potential growth of each of the main factors that contribute to sustainable growth, that is, the terms on the right-hand side of equation (16). The first term in equation (16), g_A, is the TFP growth rate. We use a growth rate of 3.3 percent for the period 1978-95 and 1.9 percent for 1995-2007, as in

Figure 1

GDP Growth, 1978-2007



Zheng, Bigsten, and Hu (2009). The second term in equation (16) is the contribution of capital (equal to the investment rate minus the depreciation rate, multiplied by an output elasticity of capital of 0.5). The third term in equation (16) is the contribution of labor: the sum of the growth rates of hours worked per person, labor force participation, and population, minus the weighted growth rate of the unemployment rate and dependency ratio. We replace the growth of hours worked per person with the growth of quality-adjusted employment (multiplied by the average years of schooling). Figure 1 shows that, starting in 2002, actual GDP growth exceeded potential growth during six consecutive years. This result is consistent with the growth accounting result based on the realized production data in Table 1.[2]

Table 1

Growth Accounting for China (percent)

Variable	1978-95	1995-2007
GDP	10.11	9.25
Capital stock	9.12	12.81
Quality-adjusted labor	3.49	2.78
TFP	3.80	1.45

SOURCE: Updated to 2007 from Zheng, Bigsten, and Hu (2009), with an output elasticity of 0.5.

Projections for the Medium Term

Many projections for China's future output potential have appeared in recent years. We provide our own estimates using the analytical framework introduced earlier. We show that it is a valid concern that China's growth pattern as measured by potential output may not be sustainable. The growth accounting result is striking when compared with what the government considers a sustainable growth target (8 percent for 2008).

Our projections rely heavily on two basic premises: (i) Capital stock growth cannot exceed GDP growth and (ii) a TFP growth rate of 2 to 3 percent must prevail for the foreseeable future. China's government was concerned about maintaining a GDP growth rate of 8 percent, both in the wake of the East Asian financial crisis of 1997 and when the Chinese economy started to overheat in 2003. We show how the "magic" 8 per-

[2] "Green" GDP estimates and TFP trends will be discussed later in the paper.

Table 2

Sustainable Growth for the Chinese Economy

α	g_K	g_L	g_A	αg_K	$(1-\alpha)g_L$	g_Y
0.5	11	3	3	5.5	1.5	10.0
0.6	11	3	3	6.6	1.2	10.8
0.5	11	3	2	5.5	1.5	9.0
0.6	11	3	2	6.6	1.2	9.8
0.5	10	3	3	5.0	1.5	9.5
0.6	10	3	3	6.0	1.2	10.2
0.5	10	3	2	5.0	1.5	8.5
0.6	10	3	2	6.0	1.2	9.2
0.5	9	3	3	4.5	1.5	9.0
0.6	9	3	3	5.4	1.2	9.6
0.5	9	3	2	4.5	1.5	8.0
0.6	9	3	2	5.4	1.2	8.6
0.5	8	3	3	4.0	1.5	8.5
0.6	8	3	3	4.8	1.2	9.0
0.5	8	3	2	4.0	1.5	7.5
0.6	8	3	2	4.8	1.2	8.0

cent growth rate can be derived from the growth accounting framework. Suppose that the borderline growth rate between extensive and intensive growth can be expressed as

$$g_Y = \frac{g_A}{(1-\alpha)} + g_L,$$

which can be derived from the usual growth accounting formula assuming that the capital stock growth rate, g_K, equals the output growth rate, g_Y. In Table 2, the GDP growth rate, g_Y, is in the far-right-hand column; other columns show combinations of parameters consistent with g_Y. With a 3 percent TFP growth rate and 0.05 output elasticity of capital, the maximum sustainable output growth rate would be 9 percent. With a 2 percent TFP growth rate and 0.06 output elasticity of capital, the maximum sustainable output growth rate would be 8 percent, which is consistent with the Chinese government's growth target for 2008 (Wen, 2008).

The magical 8 percent growth rate also has interesting implications for the structural parameters of the production function. When the assumed output elasticity of capital is 0.6, the corresponding sustainable growth rate is exactly 8 percent if TFP growth is 2 percent per year. However, 8 percent growth will not be sustainable if TFP growth is 2 percent per year but the output elasticity of capital is 0.5. Sustainable growth will be slightly more than 8 percent if TFP growth is 3 percent per year. Slower growth in the labor input (demographics) will reduce the projected output growth rate to some extent, but the trends will remain the same.

Economic growth in developing economies is considered to be mainly affected by three factors: rural-urban migration, demographics, and educational attainment. In the late 1990s, Chinese planners were preoccupied with maintaining a growth rate of 8 percent in the face of the East Asian financial crisis. Such forecasts relied on China's ability to maintain high capital forma-

tion—but if the capital growth rate exceeds the GDP growth rate, the result is extensive growth, which is likely not sustainable in the longer run, as discussed above.

We offer one more example. For simplicity, we omit the role of human capital accumulation (see Zheng, Bigsten, and Hu, 2009). Assuming, say, the output elasticity of capital is 0.5, the capital stock increases 8 percent per year, and the labor force grows slightly above 1 percent (as it has in the past decade), the TFP growth rate would be required to be 3.5 percent to achieve 8 percent GDP growth. Further, this would require the TFP contribution to GDP growth to reach 44 percent, which may be difficult to achieve in practice. Using this as a benchmark, the 5-year forecasts presented for China's 10th and 11th congressional sessions appear wildly overoptimistic because they require TFP growth to contribute 54 to 60 percent of GDP growth (see Zheng, Bigsten, and Hu, 2009).

COMPARISONS WITH THE U.S. AND EU ECONOMIES

In this section, given the structural differences, we calibrate the model to compare a typical scenario for the Chinese economy with the U.S. and EU economies. We demonstrate that growth potential varies across the three major economies because of differences in production structure, the level of development, and opportunities for absorbing foreign technologies. Growth in developed countries relies on mainly technological innovations because investment opportunities are far fewer than in developing countries. Because technology development often presents patterns of cyclical fluctuations, attempts to counterbalance business cycles or alter the trajectory of growth potentials may result in short-term gains but long-term losses. Understanding this is crucial for central banks to carry out sound monetary policies and to prevent future financial crises.

China clearly has benefitted from extensive growth in the intermediate term, but as previously shown this level of growth is not sustainable in the long run. However, China may still enjoy a high growth rate of 8 to 9 percent if it manages the transformation from extensive to intensive growth (see Table 2).

In his report to the First Session of the 11th National People's Congress in March 2008, Premier Wen Jiabao stated that "On the basis of improving the economic structure, productivity, energy efficiency and environmental protection, the GDP should grow by about eight percent" (Xinhua, 2008). This is the fourth consecutive year China set its GDP growth target at 8 percent (after five consecutive years of double-digit GDP growth) to emphasize the government's desire to promote both sound and fast economic and social development.[3] Until recently, China has tightened monetary policy to curb inflation and an overheated property market to help the transition from extensive to intensive growth.

However, it appears that China's measures to cool its economy to a sustainable level were not well timed, considering recent developments in the world economy. By November 2008, most rich economies were facing recession. The U.S. economy has been in recession since December 2007 (as confirmed by the National Bureau of Economic Research in December 2008). In November 2008, the European economy officially fell into its first recession since the euro was introduced. In China, industrial production grew by only 8.2 percent from January through October 2008, less than half the pace of the previous year and its slowest for seven years. China announced a massive stimulus package ($586 billion) in early November 2008.

Although the stimulus package was intended to boost domestic demand and create more jobs, the World Bank pointed out that the stimulus policies provide China with a good opportunity to rebalance its economy in line with the objectives of the 11th Congress's five-year plan: "The stimulus package contains many elements that support China's overall long-term development and improve people's living standards. Some of the stimulus measures give some support to the rebalancing of the pattern of growth from invest-

[3] China revised its GDP growth for 2007 from 11.9 percent to 13.0 percent and in that year overtook Germany to become the world's third-largest economy (Reuters, 2009). The growth figure was announced by the National Bureau of Statistics of China (NBSC) and was the fastest since 1993 (when GDP grew 13.5 percent).

Table 3

Growth Projections (2009-30, percent)

Countries	Capital stock	Labor	TFP	GDP
China	8.0	3.0	2.5	8.0
United States	4.0	0.5	1.2	3.1
EU	3.0	0.0	1.0	2.2

NOTE: Output elasticity of both capital and labor is 0.5 for China, 0.4 for the EU, and 0.3 for the United States.

ment, exports, and industry to consumption and services. The government can use the opportunity of the fiscal stimulus package to take more rebalancing measures, including on energy and resource pricing; health, education, and the social safety net; financial sector reform; and institutional reforms" (World Bank, 2008, p. 1).

In the longer term, China will be able to maintain its momentum as a rapidly developing economy well into the next two decades or so while the United States and EU may manage a growth rate of only 2 to 3 percent (as calibrated in Table 3).[4] Structural differences help explain the large differences in growth potential between China and the United States and EU. The contribution of capital in China is twice that in the United States. The level of development provides even greater opportunities for China than the United States and EU. Investment opportunities in China are nearly double those in the United States[5]; and the potential for China to absorb new technologies from developed nations is double that for the United States and EU.

Moreover, a shortage of labor (another important input to the production process) in developed economies will hinder faster growth of these economies in the intermediate term. In about 20 years China will face the same problem as its population ages. Demographic change due to China's baby boomers of the 1960s and 1970s entering retirement age may significantly affect the labor supply and the country's capacity to save and invest.

In the long run, economic prosperity depends on innovation-driven productivity growth. There is evidence, however, that worldwide innovations might have been ineffective in recent decades. The literature on diminishing technological opportunities since the early 1960s and recent studies on endogenous growth, which discuss related issues (see, e.g., Jones, 1999; Segerstrom, 1998; and Kortum, 1997), address this phenomenon. In a series of recent articles, Gordon (e.g., 2004) addresses the issue in terms of demand creation for new products and technological advances and suggests that the U.S. productivity revival that began in 1995 might not be sustainable (see Table 4). This suggests that the productivity slowdown that began in other developed countries in the early 1970s may continue into the next decade or so. Given the input constraints on potential output growth in the United States and EU, productivity is left as the only source of extra growth.

In this regard, historical lessons from the former Soviet Union need to be taken seriously. Soviet growth was spectacular: Its industrial

[4] The growth rate in Table 3 is somewhat too optimistic for U.S. economists: "[M]ainstream economists are exceptionally united right now around the proposition that the trend growth rate of real gross domestic product (GDP) in the United States—the rate at which the unemployment rate neither rises nor falls is in the 2 percent to 2.5 percent range" (Blinder, 2002, p. 57).

[5] Sterman (1985) presented a behavioral model of the economic long wave, which showed that "capital self-ordering" was sufficient to generate long waves. In Sterman (1983), capital self-ordering means that the capital-producing sector must order capital equipment such as large machinery to build up productive capacity. Investment expansions in the 1950s and 1960s accumulated large excess capacity in the United States and European Union. "But while stimulating basic research and training the labor force for 'new-wave' technologies are important, innovation alone will not be sufficient to lift the economy into a sustained recovery as long as excess capacity in basic industries continues to depress investment" (Sterman, 1983, p. 1276).

Table 4

Productivity Slowdowns in the Soviet Union, United States, and EU

Countries	Period	GDP	Capital	Labor	TFP
Soviet Union	1950-70	5.4	8.8	1.8	1.6
	1970-85	2.7	7.0	1.1	–0.4
United States	1950-72	3.9	2.6	1.4	1.6
	1972-96	3.3	3.1	1.7	0.6
	1996-2004	3.6	2.6	0.7	1.5
EU (euro zone)	1960-73	5.1		4.8	3.2
	1973-2003	2.2	0.5	2.8	1.0

SOURCE: Mostly period averages calculated from Ofer (1987) for the former Soviet Union, Gordon (2006) for the United States, and Musso and Westermann (2005) for the euro zone.

structure changed from an economy with an 82 percent rural population and a GNP produced mainly by agriculture to one with a 78 percent urban population and 40 to 45 percent of GNP originating in manufacturing and related industries (Ofer, 1987). This pattern of extensive growth lasted nearly 70 years, from the late 1920s to the mid-1980s. By 1970, Soviet TFP growth was zero and has been negative ever since (see Table 4). Although the current problem in Western countries is different because their patterns of growth have not been as extensive (for example, growth in capital stock has been 3 to 4 percent), their limited growth in TFP has been worrisome.

Limited TFP growth has important implications for macroeconomic planning. A straightforward strategy to boost productivity growth, of course, is to increase spending on research and development. Though many policymakers would like to believe that research and development for information and computer technologies (ICT) may benefit an economy in the long run, when managing the macroeconomy they need to consider the lag between the emergence of a new technology and the generation of sufficient demand. For example, the U.S. economy has recorded impressive productivity growth since the mid-1990s thanks to innovations and massive investments in ICT. But the ongoing financial crisis may dramatically alter the interpretation of the U.S. productivity boom of the past decade. Some critics suggest that the problem lies in the desire to maintain growth above what is sustainable by encouraging excessive investment in technology and loosening regulations for risky innovations in the financial sector. As far as macroeconomic planning is concerned, this amounts to taking the concept of "potential output" seriously.[6]

ENVIRONMENTAL CONSTRAINTS

The environment is a constraint on growth in China. Increased environmental awareness at both the central government and grassroots levels will put greater pressure on regional authorities to seek alternative patterns of growth. In Zheng, Bigsten, and Hu (2009) we note

> The Chinese government has been working on criteria and indexes of a green GDP, which deducts the cost of environmental damage and resources consumption from the traditional gross domestic product (*People's Daily*,

[6] Krugman (1997) notes that standard economic analysis suggests that the United States should not expect its economy to grow at much more than 2 percent over the next few years. He notes further that if the Federal Reserve tries to force faster growth by keeping interest rates low, serious inflation could result. Of course, inflation did not rise until recently, but the U.S. economy already started overheating in the mid-1990s. Jorgenson, Ho, and Stiroh (2006) project the best-case scenario for U.S. GDP growth to be 2.97 percent per annum for 2005-15, with an uncertainty range of 1.9 to 3.5 percent. McNamee and Magnusson (1996) give a detailed discussion on why a long-run growth rate of 2 percent could be a problem for the U.S. economy as a whole.

> March 12, 2004). Preliminary results in the recently issued *Green GDP Accounting Study Report* (2004) suggest that economic losses due to environmental pollution reached 512 billion yuan, corresponding to 3.05% of GDP in 2004, while the imputed treatment cost is 287 billion yuan, corresponding to 1.80% of GDP (The Central People's Government of the People's Republic of China, 2006). Although the concept of and measurement for green GDP are rather controversial, the report may serve as a wakeup call to the government's strategy of growth at all costs.
>
> From a productivity analysis perspective, the concept of green GDP can be straightforwardly extended to TFP, that is, green TFP. A slower green TFP growth may imply a slower (green) GDP growth. (p. 881)

We demonstrate that although the green GDP level has increased as environmental factors have been taken into account, "green TFP" growth reveals a similar trend, as shown in the main text of this article.

Environmental Factors

The World Bank (1997) first proposed the concept and calculation of "genuine domestic savings," that is, a country's saving rate calculated after subtracting from total output the costs of depletion of natural resources (especially the nonreproducible resources) and environmental pollution.

A formal model of the genuine savings rate is given by Hamilton and Clemens (1999):

$$G = GNP - C - \delta K - n(R - g) - \sigma(e - d) + m. \tag{17}$$

Here, $GNP–C$ is traditional gross savings, which includes foreign savings, where GNP is gross national product and C is consumption; $GNP–C–\delta K$ is traditional net savings, where δK is the depreciation rate of produced assets; $-n(R-g)$ is resource depletion; $S = -(R-g)$ is resource stocks, S, that grow by an amount g, are depleted by extraction R, and are assumed to be costless to produce; n is the net marginal resource rental rate; $-\sigma(e-d)$ is pollution emission costs; $X = -(e-d)$ is the growth of pollutants accumulated into a pollution stock, X, where d is the quantity of natural dissipation of the pollution stock; δ is the marginal social cost of pollution; and m is investment in human capital (current education expenditures), which does not depreciate (and may be considered as a form of disembodied knowledge).

Natural resource depletion is measured by the rent of exploiting and procuring natural resources. The rent is the difference between the producing price received by producers (measured by the international price) and total production costs, including the depreciation of fixed capital and return of capital.

Rational exploitation of natural resources is necessary for economic growth; however, if resource rents are too low, overexploitation may result. If the resources rents are not reinvested (e.g., in human resources) but instead used for consumption, the exploitation is also "irrational."

Pollution loss mostly refers to harm caused by CO_2 pollution. It is calculated by the global margin loss caused by one ton of CO_2 emissions, which Fankhauser (1995) suggests is US$20.

We expand the green GDP measure from the World Bank to include not only natural capital lost (negative factor) and education expenditure (positive factor),[7] but also net imports of primary goods (positive factor) and sanitation expenditure (positive factor). We calculate three different versions of GDP from 1978 to 2004: real GDP, World Bank–adjusted green GDP, and our author-adjusted green GDP (Table 5).

Green Capital

In the measurement of productivity, the different measures of capital formation greatly influence the measured capital stock constructed with the perpetual inventory method. We can define the green capital stock as following the method of Hamilton, Ruta, and Tajibaeva (2005):

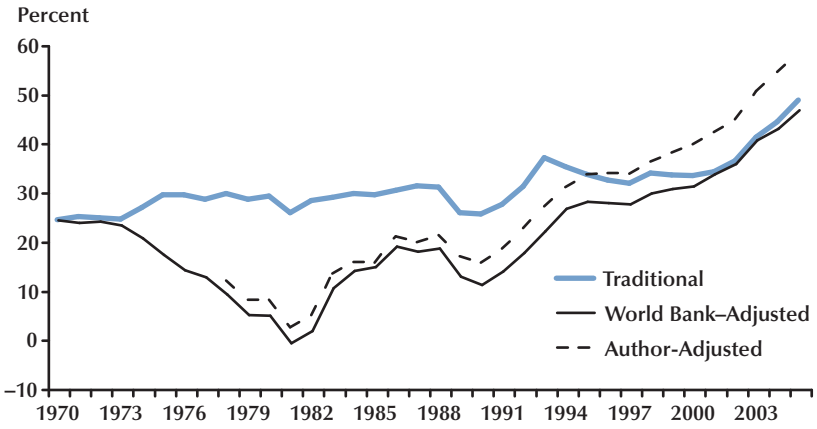
$$K'_{it} = K'_{it-1}\left(1 - \delta_{it}\right) + I', \tag{18}$$

where δ_{it} is the depreciation rate. (Time subscripts are omitted.) Our depreciation rate increases

[7] We use total education expenditures from NBSC (2006) instead of the education expenditures from World Development Indicators (2006).

Figure 2

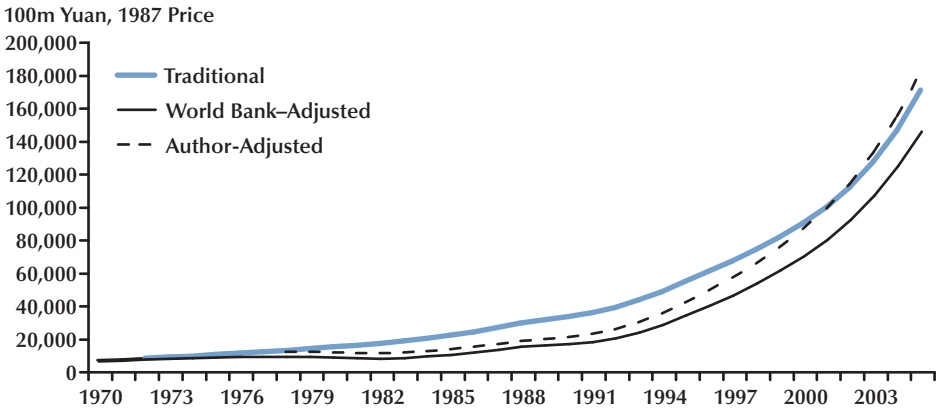
Capital Formation as a Percentage of GDP



SOURCE: NBSC (2007a) and World Bank (2006).

Figure 3

Capital Stock, 1978-2005



SOURCE: NBSC (2007a) and World Bank (2006).

Table 5

Different Measures of Green GDP (percent of real GDP)

Year	Real GDP	Natural capital lost	Education expenditure	World Bank–adjusted green GDP	Total expense on education	Total expense on sanitation	Net import of primary goods	Author-adjusted green GDP
1978	100	–23.01	1.85	78.84	2.10	3.10	–0.73	81.46
1979	100	–26.52	1.84	75.33	2.31	3.20	–0.77	78.22
1980	100	–27.54	2.08	74.54	2.51	3.30	–0.71	77.56
1981	100	–29.91	2.11	72.20	2.51	3.40	–0.77	75.23
1982	100	–28.48	2.19	73.70	2.59	3.50	–0.86	76.75
1983	100	–19.92	2.16	82.24	2.61	3.60	–1.27	85.02
1984	100	–17.35	2.07	84.72	2.51	3.30	–2.18	86.28
1985	100	–16.96	2.05	85.09	2.51	3.00	–2.80	85.75
1986	100	–12.76	2.10	89.34	2.62	3.10	–1.90	91.06
1987	100	–14.41	1.90	87.49	2.31	3.20	–1.97	89.13
1988	100	–13.57	1.87	88.30	2.22	3.30	–1.08	90.87
1989	100	–13.74	1.87	88.13	3.07	3.40	–0.74	91.99
1990	100	–15.26	1.79	86.53	3.56	4.03	–1.56	90.77
1991	100	–13.93	1.79	87.86	3.38	4.11	–1.31	92.25
1992	100	–12.50	1.70	89.20	3.25	4.09	–0.78	94.06
1993	100	–10.88	1.71	90.82	3.00	3.96	–0.40	95.68
1994	100	–8.07	2.14	94.07	3.09	3.78	–0.58	98.22
1995	100	–7.57	1.97	94.40	3.09	3.86	0.40	99.78
1996	100	–7.27	2.01	94.74	3.18	4.21	0.41	100.53
1997	100	–5.89	2.01	96.12	3.21	4.29	0.49	102.10
1998	100	–3.98	1.97	97.99	3.49	4.47	0.24	104.22
1999	100	–3.43	1.94	98.51	3.73	4.66	0.64	105.60
2000	100	–4.87	1.95	97.07	3.88	4.62	1.78	105.41
2001	100	–4.07	1.94	97.88	4.23	4.58	1.46	106.20
2002	100	–4.03	1.95	97.92	4.55	4.81	1.43	106.76
2003	100	–4.30	1.96	97.66	4.57	4.85	2.31	107.43
2004	100	–4.58	1.97	97.39	4.53	4.75	3.97	108.67

NOTE: World Bank–adjusted green GDP is the sum of columns 2, 3, and 4; the author-adjusted green GDP is the sum of columns 2, 3, 6, 7, and 8.

SOURCE: World Bank (2006) and NBSC (2006).

Table 6

GDP, Green GDP, and TFP Growth, 1978-2004 (percent)

Variable	1978-92		1992-2004		1978-2004	
GDP	9.02	(100.0)	10.12	(100.0)	9.61	(100.0)
K	7.74	(34.3)	11.27	(44.5)	9.56	(39.8)
L	2.96	(9.8)	1.07	(3.2)	2.44	(7.6)
H	2.25	(7.5)	1.90	(5.6)	2.02	(6.3)
TFP1	4.36	(48.3)	4.72	(46.6)	4.45	(46.3)
GGDP1	9.87	(100.0)	11.06	(100.0)	10.51	(100.0)
K′	5.95	(24.1)	15.88	(57.4)	10.42	(39.7)
L	2.96	(9.0)	1.07	(2.9)	2.44	(7.0)
H	2.25	(6.8)	1.90	(5.2)	2.02	(5.8)
TFP2′	5.93	(60.1)	3.82	(34.5)	5.00	(47.6)
GGDP2	10.47	(100.0)	10.75	(100.0)	10.60	(100.0)
K″	5.80	(22.2)	15.97	(59.4)	10.37	(39.1)
L	2.96	(8.5)	1.07	(3.0)	2.44	(6.9)
H	2.25	(6.4)	1.90	(5.3)	2.02	(5.7)
TFP3″	6.59	(62.9)	3.47	(32.3)	5.11	(48.2)

NOTE: *GDP* here is real GDP in 1978 prices; *GGDP1* is the World Bank–adjusted green GDP; *GGDP2* is the author-adjusted green GDP. *K* denotes capital services input, *L* labor input, *H* denotes inputs of education, sanitation expenditure, and imports of primary goods. TFP denotes total factor productivity. The shares of capital, labor, and human resource are 0.4, 0.3, and 0.3, respectively. Numbers in parentheses are the contribution ratio of each factor.

along a linear trend from 4 percent in 1952 to 6 percent in 2004. I' is the green fixed capital formation. In Figure 3, the World Bank analysis (Hamilton and Clemens, 1999) measures the green investment in any geographic or political area as

$$(19)\quad I'_{it} = I_{it} - n_{it}\left(R_{it} - g_{it}\right) - \sigma_{it}\left(e_{it} - d_{it}\right) + m_{it},$$

where I is the traditional investment, $n_{it}(R_{it} - g_{it}) - \sigma_{it}(e_{it} - d_{it})$ is the natural capital lost, and m_i is the education expenditure.

In this article, the author-adjusted green capital stock, K'', measures green investment as

$$(20)\quad \begin{aligned} I''_{it} &= I_{it} - n_{it}\left(R_{it} - g_{it}\right) \\ &-\sigma_{it}\left(e_{it} - d_{it}\right) + m_{it} + n_{it} + r_{it}, \end{aligned}$$

where m_{it} is total education expenditure (from NBSC), n_{it} is sanitation expenditure, and r_{it} is net import of primary goods.

Green TFP

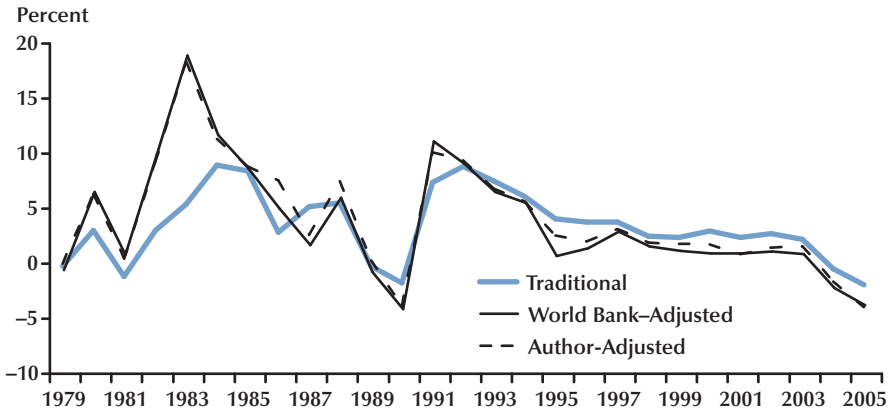
As shown in Table 6, compared with traditional GDP, the two adjusted green GDPs (the World Bank and authors' measures) have about 0.5 to 0.6 percent higher average TFP growth rates in the 1978-2004 period, with lower TFP growth in the 1992-2004 period (the author-adjusted GDP is the lowest) than in the 1978-92 period. The TFP growth rate of traditional GDP is more stable and has the opposite trend.

As shown in Figure 4, the annual TFP growth rates of the adjusted green GDPs are higher than traditional GDP in most years before 1992. They reached 13 percent higher in 1983 and then began to fall, roughly maintaining a gap of 1 to 2 percentage points with traditional GDP through 2004. In 2004, the green GDPs reached their lowest growth rate, –4 percent.

Our analysis finds that China's growth has varied between episodes of extensive and inten-

Figure 4

TFP Growth, 1979-2005



NOTE: This accounting does not include human capital. The share of capital and labor comes from Bai, Hsieh, and Qian's (2006) estimation.

sive growth. Economic growth in the 1980s was intensive growth—higher TFP growth compensated for the diminishing contribution of natural resources, that is, of "natural capital." During the 1990s, as a result of the comparative decline of its natural resource consumption, China's capital stock began to increase rapidly and its growth became more extensive, especially with respect to capital.

CONCLUSION

In this study, we have updated our previously published results on China's growth pattern, estimated China's potential output growth using official Chinese statistics, and compared China's medium-term growth perspectives with those for the United States and EU. Our findings suggest that China's extensive growth pattern might be sustainable in the intermediate term but not in the long run. However, China may still sustain a high growth rate of 8 to 9 percent if it manages the transformation from extensive to intensive growth. Several factors explain this possibility. Compared with the United States and EU, China is in a more favorable position with regard to (i) production structure, (ii) the potential to absorb new technologies, and (iii) investment opportunities. Perhaps these three factors largely explain Ed Prescott's query (1998) as to why economic "miracles" have been only a recent phenomenon.

China's reform policy since 1978 has dramatically increased its GDP as well as its role in the world economy. China was a marginal economy in 1978, but by 2007 its share of world GDP reached 5.99 percent at regular exchange rates (or 10.83 percent at purchasing power parity rates) (International Monetary Fund, 2008). This means that China now has the same economic weight as, for example, Germany. Because of China's rapid growth in recent years, its contribution to world growth has been substantial. In 2007 it was about 17 percent at regular exchange rates and as much as 33 percent at purchasing power parity rates. Even at regular exchange rates, China's contribution to global growth was considerably larger than that of the United States or EU. The global slowdown and financial contagion has now reduced the growth rate in China. Still, we believe China can continue to grow at a high rate over an extended period of time, which suggests that it will continue to be an important driver of world growth.

China's importance in the global markets for goods and services also has increased considerably. In 1978, China contributed 0.6 percent of world exports, but by 2006 its share was over 7 percent (World Bank, 2008). This is an amazingly fast expansion and entry into the global market.

The export boom China experienced during the years before the world financial crisis of 2008 could not have continued at its rapid pace even in the absence of a worldwide economic slowdown. China's rapid growth has been driven by U.S. expansionary policy, China's acceptance into the World Trade Organization, the shift of assembly plants from other countries to China, and the undervaluation of China's currency, the yuan. The impact of these factors, however, cannot be sustained. Export growth was further supported by shifting the production structure toward the international market. However, with exports approaching half of GDP, there will be less scope for further shifts. In the future it is likely that export growth will more or less keep pace with GDP growth. As long as China continues to grow faster than the world average, it will increase its global market share. China's current strategy is to shift its production toward more-sophisticated goods and, even if the impact of such is as yet limited, it is very likely that trend will continue. This means that in the future we will likely see more and more intra-industry trade between China and the Organisation for Economic Co-operation and Development (OECD).

In the short term, the world is facing an extreme international financial crisis. Debate about China's role in this crisis is intense. International financial markets are clearly more integrated than in earlier crises, although China has not opened its capital account yet. With the rapid and extended global economic growth, economic imbalances have emerged, particularly between the United States and China. The United States has been undersaving and China has been oversaving. A key issue is the character and speed of the rebalancing process. In the United States, to build savings, consumption needs to increase at a slower pace than incomes, which could hinder growth for several years. How much China can counteract this by stimulating domestic demand remains to be seen, but steps in this direction have been taken. China has implemented a large fiscal expansion. The focus of policy reforms in the near future will probably be on domestic issues, and macroeconomic policy interventions will likely seek to stimulate local demand. The process of adjustment will take a long time, however, unless there is concerted policy action.

Another important issue is how international negotiations about the future design of the financial system will evolve, particularly because of the conflict between national sovereignty and the needs of global capital markets. Still, future discussions will have to involve China in a substantial way.

China is also expanding its economic operations abroad by aggressively using its sovereign wealth to acquire assets. It currently has extensive resources, whereas other global investors are facing problems. So the current crisis is an opportunity for China to invest abroad on a larger scale than ever before. As much as 75 percent of China's investments are in developed countries to develop marketing channels, access more advanced technology, and earn a good return on its capital. There is a risk, though, that China may overpay in its eagerness to acquire assets.

China enters the current crisis better prepared than it was when hit by the 1997 East Asian financial crisis. Still, even if China today is one of the most resilient economies in the world, it may not be able to have a very large impact on the Western economies. It buys many inputs from Asia, assembles goods at home, and then sells final goods to the OECD. This means that it will suffer during a recession in wealthier countries. The export markets have also been hurt by the disruptions in the trade credit market. Most intra-Asian trade is for intermediates that are assembled in China and then exported. Few Asian exports are for Chinese demand. Thus, other Asian countries suffer when China cannot export final goods to the OECD. Countries that can supply the Chinese domestic market may be able to benefit from Chinese development, though.

Overall it is likely that China will continue to grow and increase its market share and control of wealth, which in turn will increase its economic and political influence over the longer term.

REFERENCES

Bai, Chon-gen; Hsieh, Chang-tai and Qian, Yingyi. "The Return to Capital in China." NBER Working Paper No. 12755, National Bureau of Economic Research, December 2006; www.nber.org/papers/w12755.pdf?new_window=1.

Baker, Dean. "The New Economy Does Not Lurk in the Statistical Discrepancy." *Challenge*, July-August 1998a, *41*(4), pp. 5-13.

Baker, Dean. "The Computer-Driven Productivity Boom." *Challenge*, November-December 1998b, *41*(6), pp. 5-8.

Baker, Dean. "The Supply-Side Effect of a Stock Market Crash." *Challenge*, September-October 2000, *43*(5), pp. 107-17.

Baker, Dean. "Is the New Economy Wearing Out?" *Challenge*, January-February 2002, *45*(1), pp. 117-21.

Blinder, Alan S. "The Speed Limit: Fact and Fancy in the Growth Debate." *American Prospect*, September-October 1997, Issue 34, pp. 57-62.

Bosworth, Barry P. and Triplett, Jack E. "The Early 21st Century Productivity Expansion Is Still in Services." *International Productivity Monitor*, Spring 2007, Issue 14, pp. 3-19.

Brenner, Robert. "The Boom and the Bubble." *New Left Review*, November-December 2000, *6*, pp. 5-39.

Brenner, Robert. "New Boom or New Bubble?" *New Left Review*, January-February 2004, *25*, pp. 57-100.

Central People's Government of the People's Republic of China. *Green GDP Accounting Study Report 2004*. September 11, 2006; www.gov.cn/english/2006-09/11/content_384596.htm.

China Cement. "Discussion of Financing Mode for Improving the Energy Efficiency of China's Cement Industry" (in Chinese). 2007.

Chow, Gregory C. "Another Look at the Rate of Increase in TFP in China." *Journal of Chinese Economic and Business Studies*, May 2008, *6*(2), pp. 219-24.

Chow, Gregory C. and Li, Kui-Wai. "China's Economic Growth: 1952-2010." *Economic Development and Cultural Change*, October 2002, *51*(1), pp. 247-56.

Congressional Budget Office. "CBO's Method for Estimating Potential Output: An Update." August 2001; www.cbo.gov/ftpdocs/30xx/doc3020/PotentialOutput.pdf.

Congressional Budget Office. "R&D and Productivity Growth: A Background Paper." June 2005.

Fankhauser, Samuel. *Valuing Climate Change: The Economics of the Greenhouse*. London: Earthscan, 1995.

Gordon, Robert J. "Does the New Economy Measure Up to the Great Inventions of the Past?" *Journal of Economic Perspectives*, Fall 2000, *14*(4), pp. 49-74.

Gordon, Robert J. "Hi-tech Innovation and Productivity Growth: Does Supply Create Its Own Demand?" NBER Working Paper No. 9437, National Bureau of Economic Research, January 2003; www.nber.org/papers/w9437.pdf?new_window=1.

Gordon, Robert J. "Five Puzzles in the Behaviour of Productivity, Investment and Innovation." CEPR Discussion Paper No. 4414, Centre for Economic and Policy Research, June 2004.

Gordon, Robert J. "The 1920s and the 1990s in Mutual Reflection." NBER Working Paper No. W11778, National Bureau of Economic Research, November 2005; www.nber.org/papers/w11778.pdf?new_window=1.

Gordon, Robert J. "Future U.S. Productivity Growth: Looking Ahead by Looking Back." Presented at the *Workshop at the Occasion of Angus Maddison's 80th Birthday, World Economic Performance: Past, Present, and Future*. University of Groningen, The Netherlands, October 27, 2006.

Hamilton, Kirk and Clemens, Michael. "Genuine Savings Rates in Developing Countries." Environment Department, World Bank, 1998.

Hamilton, Kirk; Ruta, Giovanni and Tajibaeva, Liaila. "Capital Accumulation and Resource Depletion:

A Hartwick Rule Counterfactual." Policy Research Working Paper 3480, World Bank, January 2005.

Holz, Carsten A. "The Quantity and Quality of Labor in China 1978-2000-2025." Working Paper, Hong Kong University of Science and Technology, May 2005; http://ihome.ust.hk/~socholz/Labor/Holz-Labor-quantity-quality-2July05-web.pdf.

International Monetary Fund. *World Economic Outlook: Financial Stress, Downturns, and Recoveries*. Washington, DC: IMF, October 2008.

Irmen, Andreas. "Extensive and Intensive Growth in a Neoclassical Framework." *Journal of Economic Dynamics and Control*, August 2005, *29*(8), pp. 1427-48.

Jones, Charles I. "Growth: With or Without Scale Effects?" *American Economic Review*, May 1999, *89*(2), pp. 139-44.

Jorgenson, Dale W.; Ho, Mun S.; Samuels, Jon D. and Stiroh, Kevin J. "Industry Origins of the American Productivity Resurgence." *Economic Systems Research*, September 2007, *19*(3), pp. 229-52.

Jorgenson, Dale W.; Ho, Mun S. and Stiroh, Kevin J. "Potential Growth of the U.S. Economy: Will the Productivity Resurgence Continue?" *Business Economics*, January 2006, *41*(1), pp. 7-16.

Kim, Jong-Il and Lau, Lawrence J. "The Sources of Economic Growth of the East Asian Newly Industrialized Countries." *Journal of the Japanese and International Economies*, September 1994, *8*(3), pp. 235-71.

Kortum, Samuel S. "Research, Patenting, and Technological Change." *Econometrica*, November 1997, *65*(6), pp. 1389-419.

Krugman, Paul. "The Myth of Asia's Miracle." *Foreign Affairs*, November-December 1994, *73*(6), pp. 62-78.

Krugman, Paul. "How Fast Can the U.S. Economy Grow?" *Harvard Business Review*, July-August 1997, *75*(4), pp. 123-29.

Kuijs, Louis and Wang, Tao. "China's Pattern of Growth: Moving to Sustainability and Reducing Inequality." *China and World Economy*, January-February 2006, *14*(1), pp. 1-14.

Lin, Justin Yifu; Cai, Fang and Li, Zhou. *The China Miracle: Development Strategy and Economic Reform*. Hong Kong: Chinese University Press, 1996.

McNamee, Mike and Magnusson, Paul. "Let's Get Growing: The Economy Can Run Faster. Here's How To Make It Happen." *Business Week*, July 8, 1996, p. 90-98.

Musso, Alberto and Westermann, Thomas. "Assessing Potential Output Growth in the Euro Area: A Growth Accounting Perspective." ECB Occasional Paper No. 22, European Central Bank, January 2005; www.ecb.int/pub/pdf/scpops/ecbocp22.pdf.

Nan, Liangjin and Xue, Jinjun. "Estimation of China's Population and Labor Force, 1949-1999" (in Chinese). *China Population Science*, 2002, No. 4, pp. 1-16.

National Bureau of Statistics of China. *China Statistical Yearbook*. Beijing: China Statistics Press, 2005a, 2006, 2007a, and 2008.

National Bureau of Statistics of China. *Comprehensive Statistical Data and Materials on 55 Years of New China*. Beijing: China Statistics Press, 2005b.

National Bureau of Statistics of China. *China Labour Statistical Yearbook*. Beijing: China Statistics Press, 2007b.

National Bureau of Statistics of China. *Historical Data on China's Gross Domestic Product Accounting, 1952-2004*. Beijing: China Statistics Press, 2007c.

Nelson, Richard R. and Romer, Paul M. "Science, Economic Growth, and Public Policy." *Challenge*, March-April 1996, *39*(2), pp. 9-21.

Ofer, Gur. "Soviet Economic Growth: 1928-1985." *Journal of Economic Literature*, December 1987, *25*(4), pp. 1767-833.

Oliner, Stephen D.; Sichel, Daniel E. and Stiroh, Kevin J. "Explaining a Productive Decade." *Brookings Papers on Economic Activity*, 2007, *1*, pp. 81-152.

People's Daily (Beijing, China). "Green GDP System to Debut in 3-5 Years in China." March 12, 2004.

Phelps, Edmund S. "The Boom and the Slump: A Causal Account of the 1990s/2000s and the 1920s/1930s." *Journal of Policy Reform*, March 2004, *7*(1), pp. 3-19.

Prescott, Edward C. "Needed: A Theory of Total Factor Productivity." *International Economic Review*, August 1998, *39*(3), pp. 525-51.

Prescott, Edward C. "Richard T. Ely Lecture: Prosperity and Depression."*American Economic Review*, May 2002, *92*(2), pp. 1-15.

Proietti, Tommaso; Musso, Alberto and Westermann, Thomas. "Estimating Potential Output and the Output Gap for the Euro Area: A Model-Based Production Function Approach." *Empirical Economics*, July 2007, *33*(1), pp. 85-113.

Reuters. "China's Revised 2007 GDP Growth Moves It Past Germany." January 15, 2009.

Romer, David. *Advanced Macroeconomics*. Third Edition. Boston: McGraw-Hill Irwin, 2006.

Schiff, Lenore. "Economic Intelligence: Is the Long Wave About To Turn Up?" *Fortune*, February 22, 1993, *127*(4), p. 24.

Segerstrom, Paul S. "Endogenous Growth without Scale Effects." *American Economic Review*, December 1998, *88*(5), pp. 1290-310.

Sterman, John D. "The Long Wave." *Science*, March 18, 1983, *219*(4590), p. 1276.

Sterman, John D. "A Behavioral Model of the Economic Long Wave." *Journal of Economic Behavior and Organization*, 1985, *6*(1), pp. 17-53.

Sterman, John. "The Long Wave Decline and the Politics of Depression." *Bank Credit Analyst*, 1992, *44*(4), pp. 26-42.

Solow, Robert M. "Perspectives on Growth Theory." *Journal of Economic Perspectives*, Winter 1994, *8*(1), pp. 45-54.

Stiglitz, Joseph. "The Roaring Nineties." *Atlantic Monthly*, October 2002, *290*(3), pp. 76-89.

Stiroh, Kevin J. "Is There a New Economy?" *Challenge*, July-August 1999, *42*(4), pp. 82-101.

Vatter, Harold G. and Walker, John F. "Did the 1990s Inaugurate a New Economy?" *Challenge*, January-February 2001, *44*(1), pp. 90-116.

Walsh, John. "Is R&D the Key to the Productivity Problem?" *Science*, February 1981, *211*(13), pp. 685-88.

Wen, Jiabao. "Report on the Work of the Government." Presented at the First Session of the 11th National People's Congress, March 5, 2008.

World Bank. *Expanding the Measure of Wealth: Indicators of Environmentally Sustainable Development*. Washington, DC: Environment Department, World Bank, 1997.

World Bank. *World Development Indicator CD-ROM*. Washington, DC: World Bank, 2006 and 2008.

World Bank. *China Quarterly Update*. Washington, DC: World Bank, December 2008.

Xinhua. "Highlights of Chinese Premier Wen Jiabao's Government Work Report." March 8, 2008, update.

Young, Alwyn. "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience." *Quarterly Journal of Economics*, August 1995, *110*(3), pp. 641-80.

Zheng, Jinghai. "On Chinese Productivity Studies." *Journal of Chinese Economic and Business Studies*, May 2008, *6*(2, Special Issue), pp. 109-19.

Zheng, Jinghai; Bigsten, Arne and Hu, Angang. "Can China's Growth Be Sustained: A Productivity Perspective?" *World Development*, April 2009, *37*(4, Special Issue), pp. 874-88.

Zheng, Jinghai and Hu, Angang. "An Empirical Analysis of Provincial Productivity in China (1979-2001)." *Journal of Chinese Economic and Business Studies*, 2006, *4*(3), pp. 221-39.

APPENDIX: DATA DESCRIPTION

The main variables investigated in the study are aggregate output (GDP at a constant price), aggregate labor (the number of people employed), and capital stock (accumulated fixed capital investment at a constant price). For details of the treatment of data, see Zheng, Bigsten, and Hu (2009, appendix). Here we outline the data used in addition to those in that study.

Capital Stock

We have collected a series of capital stock, which is the accumulation of total social fixed asset investment since 1978. We use the price indices of gross fixed capital formation from *Historical Data of China's Gross Domestic Product Accounting, 1952-2004* (NBSC, 2007c) to deflate investment data before 1990. For investment after 1990, we use the price indices of fixed asset investment from the *China Statistical Yearbook* (NBSC, 2005a, 2006a, 2007a, and 2008) See Figures A1 through A3 for time plots of the series and related measures.

Labor

The labor force data used are the economically active population data from *Comprehensive Statistical Data and Materials on 55 Years of New China* (NBSC, 2005b) and are extended to 2007 based on the growth rate for each year from the *China Statistical Yearbook* (NBSC, 2005a, 2006, 2007a, and 2008). Because of the inconsistency of official data before 1990, we use an adjusted series of labor force from Nan and Xue (2002) to update our pre-1990 data. The data on employment are from the *China Labour Statistical Yearbook* (NBSC, 2007b). We generate a new series of the data before 1990 based on the official unemployment (defined as the gap of economic active population and employment) and the labor force data from Nan and Xue (2002). (See Figures A4 to A6.)

Human Capital

To measure human capital, we use average years of schooling of Chinese laborers to adjust for labor quality improvement. Data for 1978-2005 are from Holz (2005) and include two series, one with and one without military service included. We use the former. "Labor" is defined as quality-adjusted laborers, that is, the number of employees multiplied by the average years of schooling. (See Figures A7 to A9.)

Energy Consumption and Carbon Dioxide Emission

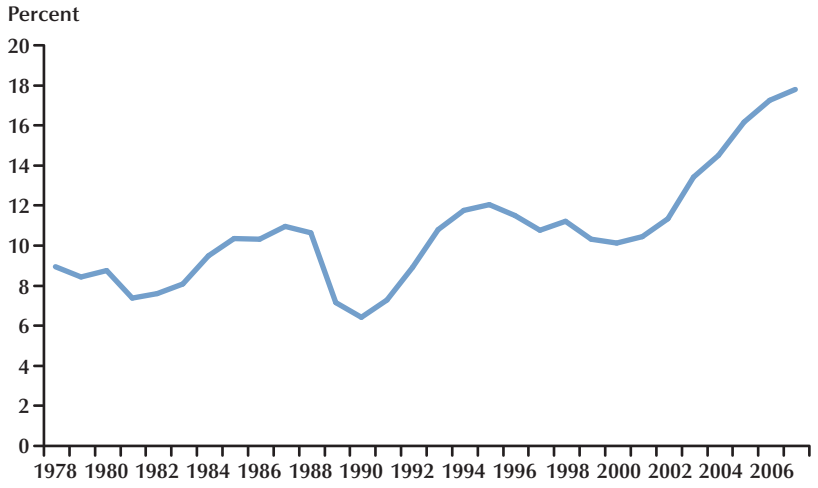
Energy consumption data and its structure are from *Comprehensive Statistical Data and Materials on 55 Years of New China* (NBSC, 2005b), which provides consumption of fossil fuel (to estimate the carbon dioxide [CO_2] emissions) together with cement production data. CO_2 emissions based on energy consumption is calculated according to the following formula:

CO_2 Emissions = Consumption of Fossil Fuel[8] × Carbon Emission Factor × Fraction of Carbon Oxidized + Production of Cement × Processing Emission Factor.

The fraction of carbon oxidized refers to the ratio of carbon oxidized to the quantity of CO_2 emitted, which is a constant ratio 3.67 (44:12). The most important coefficient here is the carbon emission factor, which refers to the equivalent carbon emissions in the consumption of fossil fuel. We use the factor from the Energy Research Institute of China's National Development and Reform Committee, which is 0.67 per ton of coal-equivalent fuel. Further, the production of cement emits more CO_2 than the consumption of fossil fuel because of the calcining of limestone, which on average creates 0.365 tons of CO_2 per ton of cement produced (China Cement, 2007).

Figure A1

Gross Capital Stock Growth



[8] A more-accurate calculation would exclude the carbon sink. We use the approximate amount because of the limited availablity of data.

Figure A2

Gross Capital Stock and Its Components

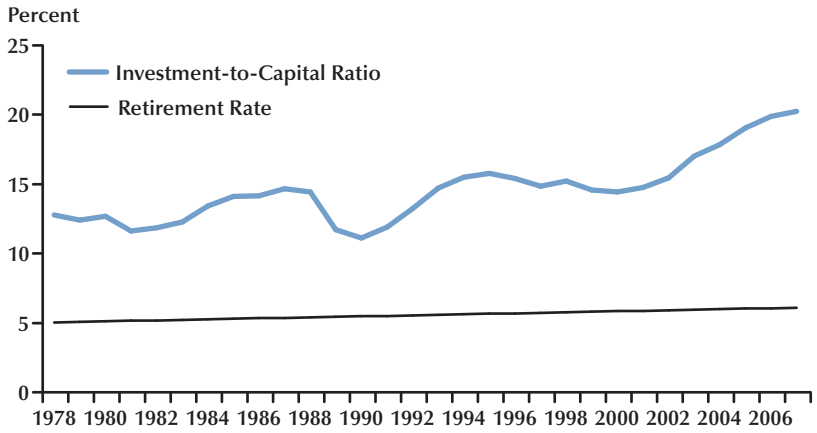


Figure A3

Determinants of the Investment-to-Capital Ratio

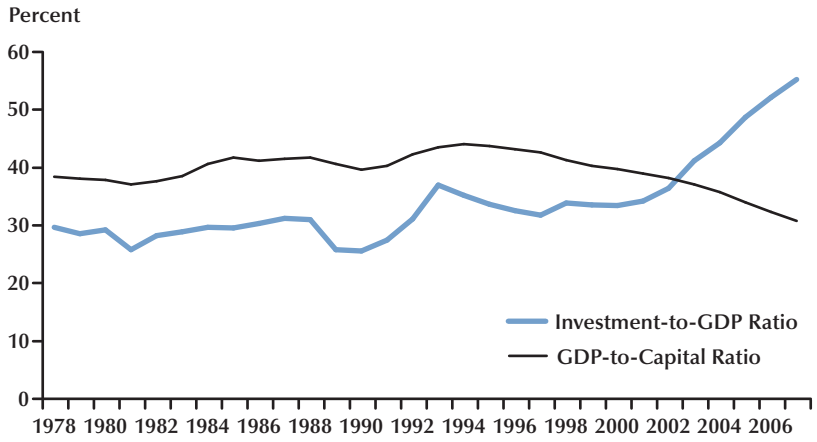


Figure A4

Labor Force Growth

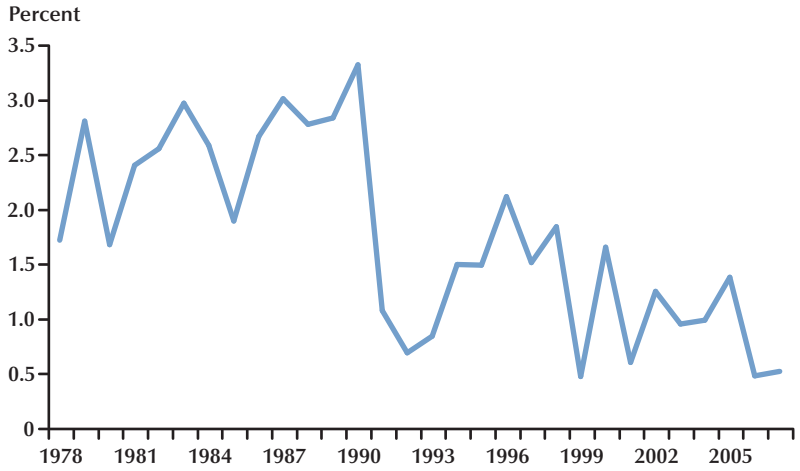


Figure A5

Unemployment Rate

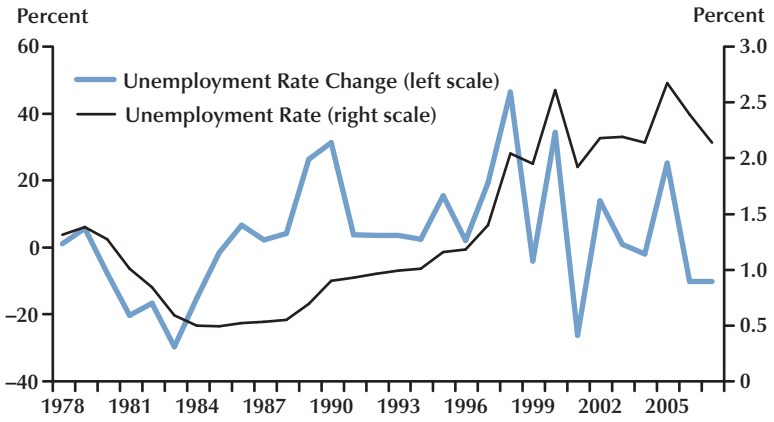


Figure A6

Working-Age Population Growth

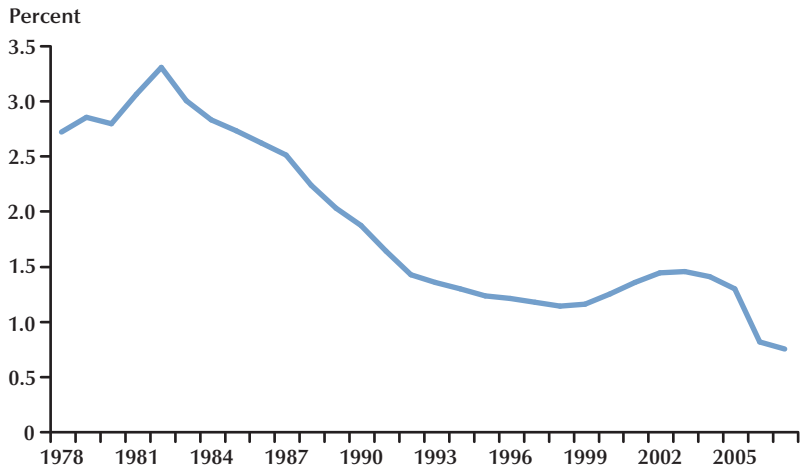


Figure A7

Participation Rate

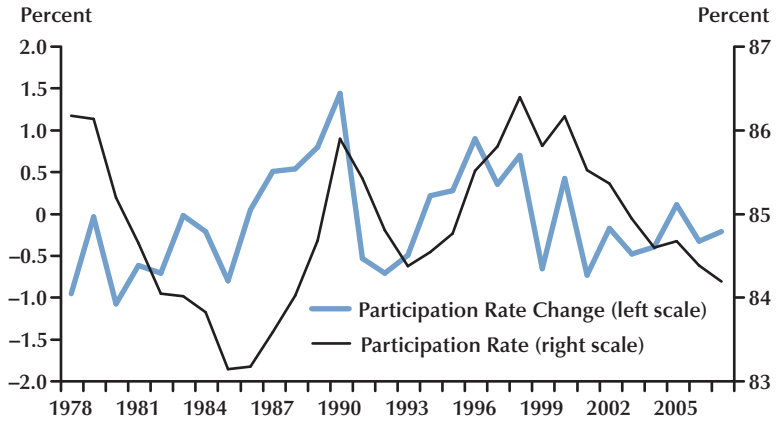


Figure A8

Population Growth

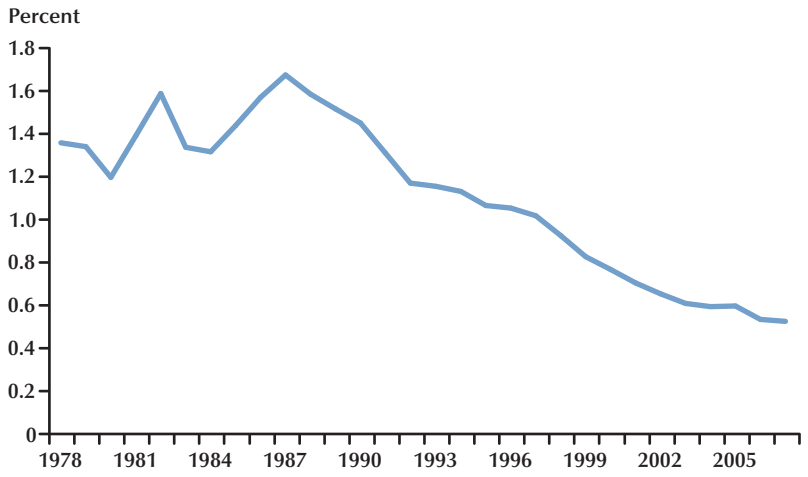
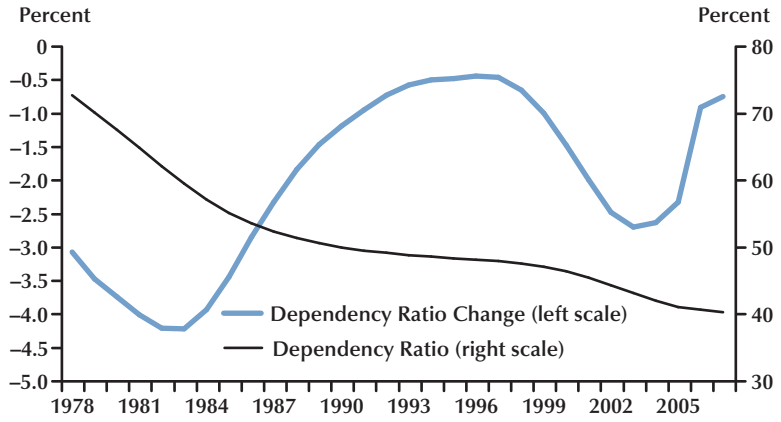


Figure A9

Dependency Ratio



Commentary

Xiaodong Zhu

China's growth performance over the past three decades has been remarkable, if not unprecedented. A natural question is whether China's recent pattern of growth is sustainable in the long run. Zheng, Hu, and Bigsten (2009) use a standard growth accounting framework to address this question. They assume that the aggregate production function is Cobb-Douglas:

$$Y_t = A_t K_t^{1-\alpha} L_t^{\alpha},$$

where A_t, K_t, and L_t are total factor productivity (TFP), capital stock, and employment, respectively, and α is the income share of labor. According to their calculation using a labor share of 0.5, the contribution of TFP growth to China's gross domestic product (GDP) growth has declined in recent years. As they reported in their Table 1, the average annual growth rates of GDP and TFP were 10.11 percent and 3.8 percent, respectively, for 1978-95 but 9.25 percent and 1.45 percent, respectively, for 1995-2007. In other words, the contribution of TFP growth to GDP growth declined by 38 percent in the first period and 16 percent in the second period. In contrast, the average growth rate of the capital stock increased from 9.12 percent in the first period to 12.81 percent in the second period. So the contribution of physical capital accumulation increased from 45 percent in the first period to 69 percent in the second period. Based on these calculations, the authors suggest that in recent years China has pursued an *extensive* growth strategy that relies heavily on capital accumulation rather than TFP growth. Because investment as a percentage of GDP has exceeded 40 percent, the authors argue that further increases in the investment rate, which would be needed to maintain a growth rate of capital stock similar to its recent average, is not sustainable and therefore extensive growth cannot be sustained in the long run. They suggest that a switch from *extensive* to *intensive* growth is needed for China to sustain its recent growth performance, thus the emphasis on productivity increases.

The paper addresses an important question, and growth accounting is the right place to start. I am also sympathetic to the authors' arguments, especially their suggestion that TFP growth is crucial for China's growth performance in the long run. However, a few puzzling facts about China's recent growth performance need to be accounted for before we can judge the relative role of capital accumulation and TFP in China's recent growth and make projections about its future growth.

First, given the high investment rates in recent years, low returns to capital might be expected. However, Bai, Hsieh, and Qian (2006) show that this is not the case. They find that China's returns to capital have been around 20 percent in recent years, which is not significantly lower than returns to capital worldwide. If there has been no significant TFP growth, how could China increase its investment rate without lowering the returns to capital?

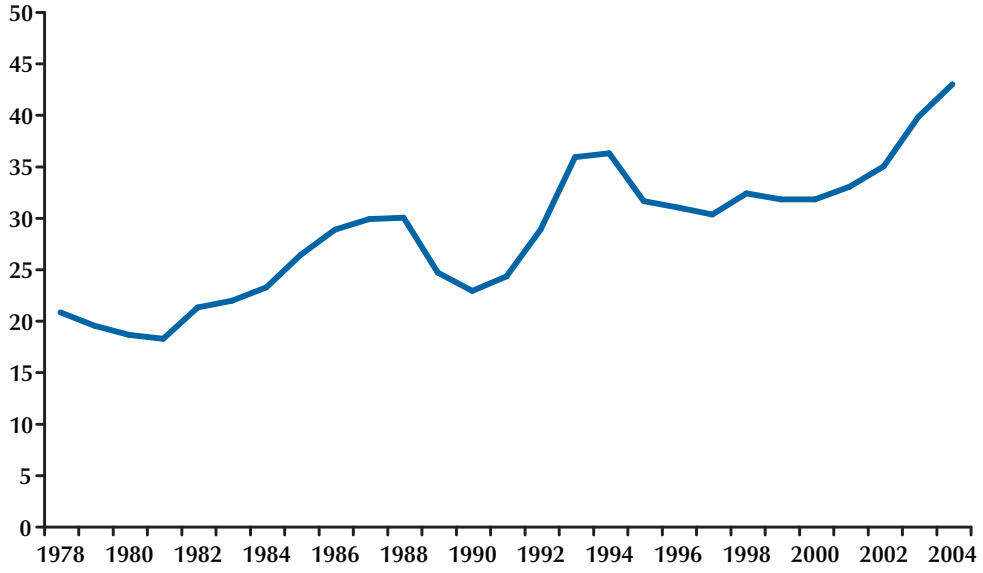
Second, since 1978, when economic reform started in China, TFP has grown substantially.

Xiaodong Zhu is a professor of economics at the University of Toronto and a special term professor at Tsinghua University in Beijing.

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 343-47.
© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

Figure 1

China's Investment-to-GDP Ratio



SOURCE: Brandt and Zhu (2009).

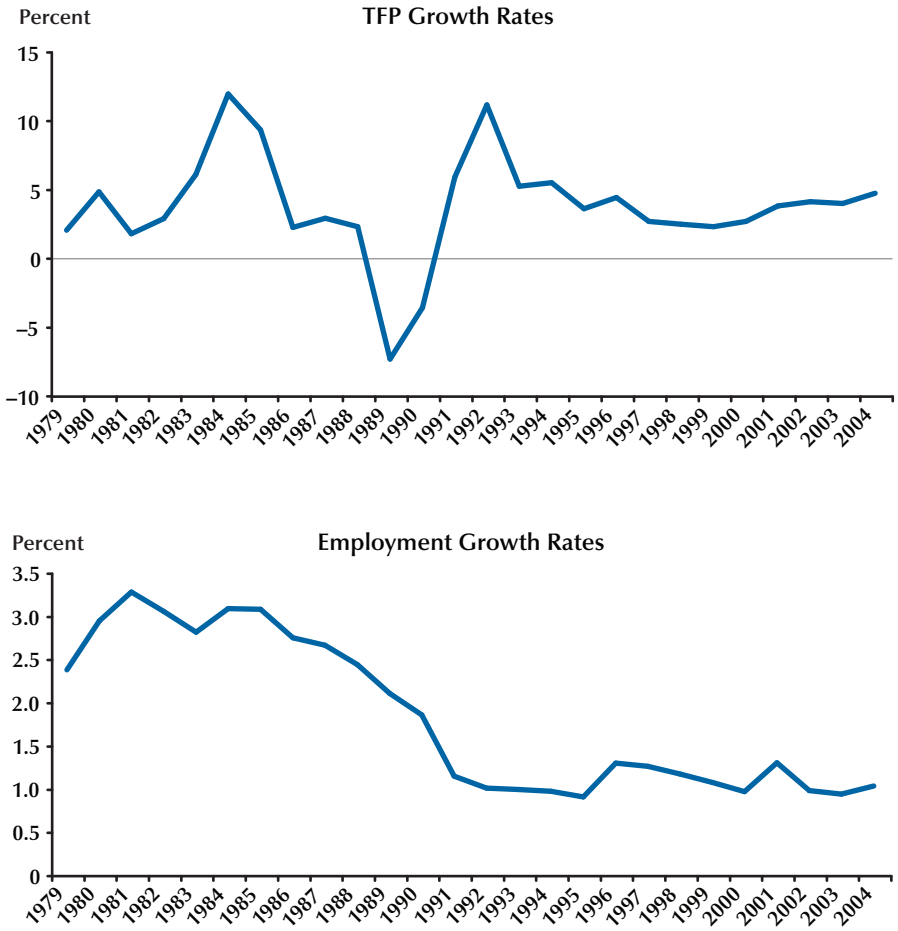
According to a standard neoclassical growth model, an increase in the TFP growth rate would result in a sharp and immediate increase in the investment rate followed by a gradual decline. The actual investment rate in China, however, behaves quite differently. Figure 1 shows that it has increased gradually over time. Arguably, this gradual increase in the investment rate may have been due to a gradual increase in the growth rate of TFP or the labor input. Figure 2 shows the growth rates of TFP and employment in China and neither has had an upward trend. Why, then, didn't the investment rate grow more rapidly?

The answers to these questions are important for understanding the nature of China's growth performance and cannot be easily answered using an aggregate growth accounting framework. I suggest addressing these questions by looking at more disaggregated data. Figure 3 shows the returns-to-capital and capital-to-labor ratios in the state and non-state nonagricultural sectors, respectively, and their significant differences. In the state sector, the capital-to-labor ratio increased steadily before 1997 and dramatically afterward. Correspondingly, returns to capital were roughly constant at 10 percent before 1997 and declined sharply afterward. Such behavior is consistent with what Zheng, Hu, and Bigsten (2009) find at the aggregate level. It suggests that, in the state sector, capital accumulation played a much more important role than TFP growth in recent years. For the non-state sector, however, the story is quite different. The capital-to-labor ratio in this sector actually declined in the early years, which coupled with TFP growth resulted in a sharp increase in returns to capital. In recent years, the non-state sector's capital-to-labor ratio increased, but the returns to capital did not decline. This sector has maintained a relatively high rate of returns to capital (around 60 percent) because of rapid TFP growth (Figure 4).

So, the answer to the question of whether China's recent growth pattern is extensive or intensive depends on which part of the Chinese economy is analyzed. If the focus is on the state sector, then it clearly follows an extensive growth path.

Figure 2

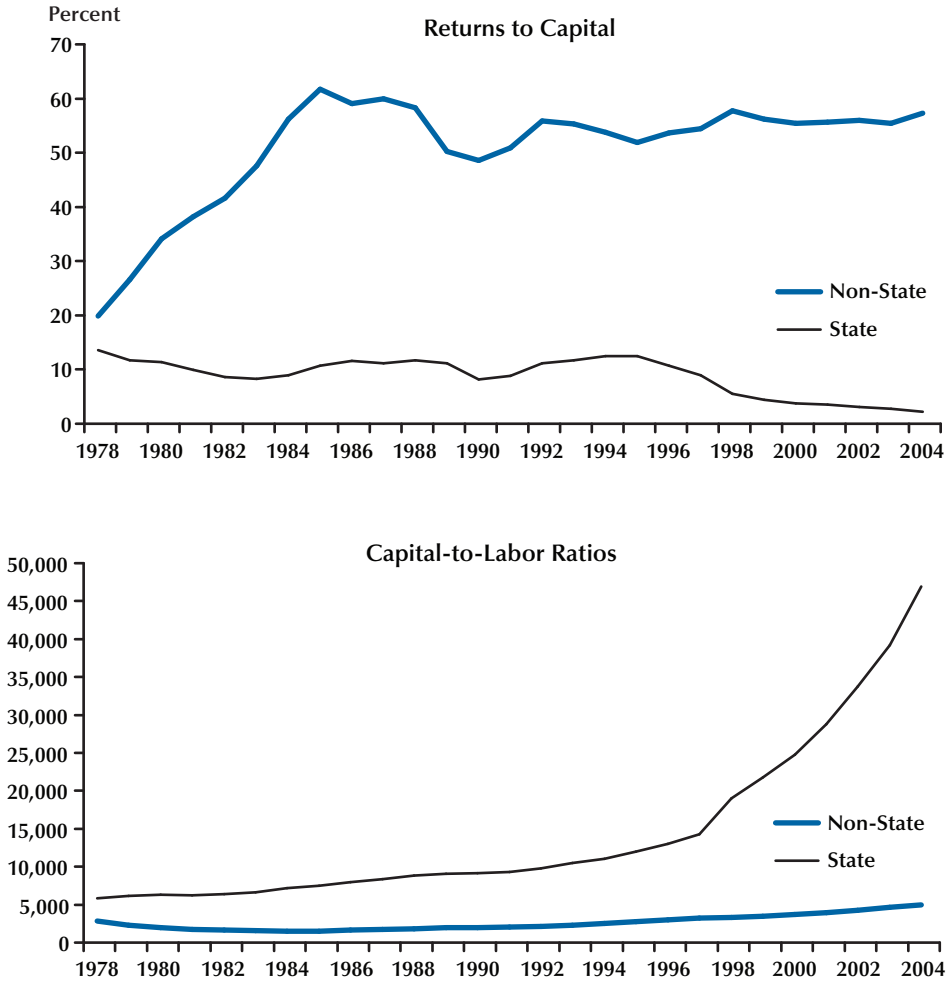
China's TFP and Employment Growth Rates



SOURCE: Brandt and Zhu (2009).

Figure 3

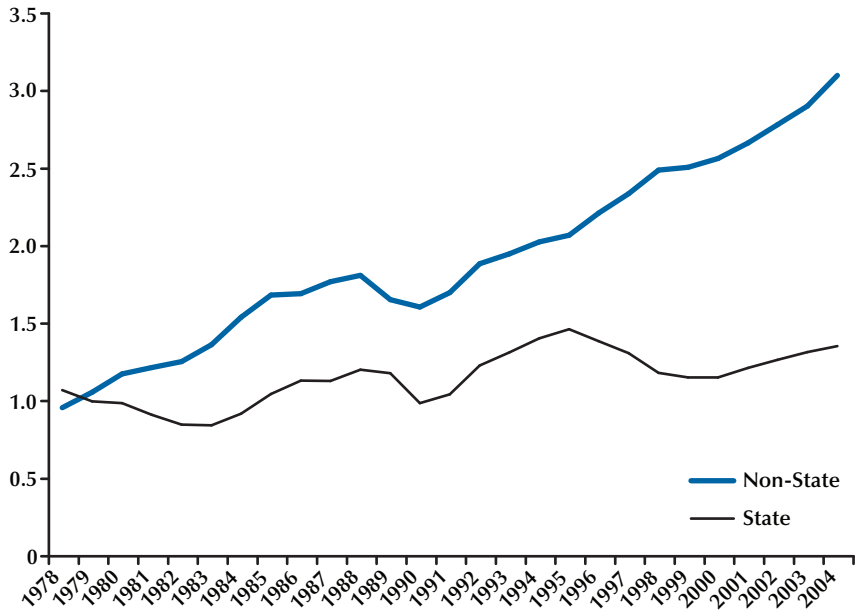
China's Returns to Capital and Capital-to-Labor Ratios



SOURCE: Brandt and Zhu (2009).

Figure 4

China's TFP



SOURCE: Brandt and Zhu (2009).

The non-state sector, on the other hand, follows an intensive growth path that relies much more on TFP growth than capital accumulation. As Zheng, Hu, and Bigsten argue in their paper, intensive growth is more likely to be sustainable than extensive growth. The sustainability of China's recent growth performance, then, will depend on the relative importance of the two sectors. Measured by the share of employment, the non-state sector's importance has increased over time. According to Brandt and Zhu's (2009) estimates, the non-state sector's share of nonagricultural employment increased from 48 percent in 1978 to 87 percent in 2004. Measured by the share of investment, however, the picture of the non-state sector is not as rosy.

Despite its lackluster TFP growth performance and declining employment share, the state sector's share of investment has always stayed above 60 percent. Given the high TFP growth in the non-state sector and the high investment rate in the state sector, China can increase both the aggregate efficiency of the economy and the GDP growth rate without increasing the aggregate investment rate, by shifting investment from the state sector to the non-state sector.

REFERENCES

Bai, Chong-En; Hsieh, Chang-Tai and Qian, Yingyi. "The Return to Capital in China." *Brookings Paper on Economic Activity*, 2006, Issue 2, pp. 61-88.

Brandt, Loren and Zhu, Xiaodong. "Explaining China's Growth." Working paper, University of Toronto, 2009.

Zheng, Jinghai; Hu, Angang and Bigsten, Arne. "Potential Output in a Rapidly Developing Economy: A Comparison of China with the United States and the European Union." Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 317-42.

Estimating U.S. Output Growth with Vintage Data in a State-Space Framework

Richard G. Anderson and Charles S. Gascon

This study uses a state-space model to estimate the "true" unobserved measure of total output in the U.S. economy. The analysis uses the entire history (i.e., all vintages) of selected real-time data series to compute revisions and corresponding statistics for those series. The revision statistics, along with the most recent data vintage, are used in a state-space model to extract filtered estimates of the "true" series. Under certain assumptions, Monte Carlo simulations suggest this framework can improve published estimates by as much as 30 percent, lasting an average of 11 periods. Real-time experiments using a measure of real gross domestic product show improvement closer to 10 percent, lasting for 1 to 2 quarters. (JEL C10, C53, E01)

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 349-69.

Statistical agencies face a tradeoff between accuracy and timely reporting of macroeconomic data. As a result, agencies release their best estimates of the "true" unobserved series in the proceeding month, quarter, or year with some measurement error.[1] As agencies collect more information, they revise their estimates, and the data are said to be more "mature." As the reported data mature, the estimates, on average, are assumed to converge toward the "true" unobserved values. This study examines a methodology in which the "true" value of an economic variable is latent in the sense of the state vector in a state-space model. In doing so, we use recent modeling suggestions by Jacobs and van Norden (2006) and Cunningham et al. (2007) regarding relationships among real-time data, measurement error as a heteroskedastic stochastic process, and the latent, "true" data for an economic variable of interest.

The importance of *potential output growth* in policymaking motivates our study. Forward-looking macroeconomic models suggest that the predicted future path of the output gap should be important to policymakers. To the extent that policymakers are concerned with a Federal Reserve–style "dual mandate," an output gap equal to 1 percent of potential output may be quite alarming if projections suggest it will continue, but relatively innocuous if the gap is expected to shrink rapidly during the next few quarters. Recent studies on inflation forecasting conclude that the output gap, when measured in real-time using vintage data, has little predictive power for inflation (e.g., Orphanides and van Norden, 2005; and Stock and Watson, 2007). It is also important to study the real-time measurement of potential output because policymakers occasionally face

[1] In Appendix B we address the philosophical question of why an econometrician might believe s/he can improve published data—after all, statisticians who produce data have access to the same historical data used by econometricians and, hence, should create models using the same understanding of the revision process that econometricians use. Over long periods, benchmarks and redefinitions muddy the analysis. But, it is an act of hubris to assert that any simple statistical model can produce consistently more-accurate near-term data than are produced by the specialists constructing the published data. Hubris aside, we have written this paper regardless.

Richard G. Anderson is a vice president and economist and Charles S. Gascon is a senior research associate at the Federal Reserve Bank of St. Louis.

© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

possible changes/breaks in the underlying growth trend of productivity and, hence, potential output.

Our objective in this study is not to assess inflation-forecasting models, although that has been a major use of potential output measures; rather, it is to estimate the "true" value of real output for use in the construction of trend-like measures of potential output. One of the larger recent studies in this vein, albeit focused on inflation prediction, is by Orphanides and van Norden (2005). The study considers, as predictive variables for inflation, both a wide range of output gap measures (which differ with respect to data vintage and the trend estimator) and lagged values of real output growth. Their conclusion regarding output gap models as predictors of inflation is straightforward—the output gap does not reliably predict inflation, although the differences in forecast performance between output-gap and output-growth models are not statistically significant:

> [O]ur analysis suggests that a practitioner could do well by simply taking into account the information contained in real output growth without attempting to measure the level of the output gap. This model was consistently among the best performers, particularly over the post-1983 forecast sample. (p. 597)

Motivated by these findings, this article models the true (unobserved) output measure of real output and the implications of such for estimators of a real output trend. To so do, we explore the measurement error and subsequent data-revision process for real gross domestic product (RGDP).

LITERATURE REVIEW

Early studies of real-time data focused on the sensitivity of certain statistics to data vintage (Howrey, 1978 and 1984; Croushore and Stark, 2001; Diebold and Rudebusch, 1991; and Orphanides and van Norden, 2002 and 2005). Later research posed the problem more formally as a signal-extraction problem (Kishor and Koenig, 2005; Aruoba, Diebold, and Scotti, 2008; and Aruoba, 2008). Both approaches emphasized the sensitivity of statistical inferences, including measures of the forecasting power of the output gap.

Recent analyses have focused on "the possibility that the sequence of vintages released over time may contain useful information with which to interpret the most recent vintage of data and to anticipate future outcomes" (Garratt et al., 2008, p. 792). Such a possibility was discussed by Howrey (1978) but only recently has become the centerpiece of certain studies.

A long literature has addressed the use of real-time data, starting with Howrey's 1978 paper on forecasting with preliminary data and including Croushore and Stark's (2000) release of a vintage economic dataset at the Philadelphia Fed. This literature, until recently, has focused on three main issues: (i) embedding an estimate of the data revision process into forecasting models, (ii) assessing the sensitivity of statistical inferences in macroeconomic data to data vintage, and (iii) checking the forecastability of revisions, in the context of Mankiw and Shapiro's (1986) classic discussion of "news vs. noise."[2]

Some authors have argued there are policy implications of such issues. Croushore (2007) argues that revisions to published personal consumption expenditures (PCE) inflation rates are forecastable, at least from August to August of the following year, and identifies an upward bias to revisions, indicating that initial estimates consistently are too low. He suggests that policymakers should "account for" this bias and predictability in setting monetary policy. Kozicki (2004) analyzes vintages of the output gap, employment gap, and inflation data and finds that revised data and real-time data suggest differing policy actions. Kozicki suggests that policymakers should place greater emphasis on more-certain data and be less aggressive in response to changes in data subject to large revisions. Previously, Orphanides and van Norden (2002 and 2005) argued that failure to appreciate the difference between real-time and final data risks serious policy errors.

[2] Our analysis is silent on the discussion of "news vs. noise" in real-time data analysis—"news" meaning that the statistical agency publishes efficient estimates using all available information, "noise" meaning there is measurement error unrelated to the true value. These are not mutually exclusive; both conditions may not hold. News implies revisions have mean zero, noise does not. Empirical results suggest that noise dominates the data-generating process.

Recently, data availability has encouraged researchers to explore a methodology in which they estimate the "true" values and measurement errors as a latent state vector. In such studies, the revisions are modeled as a statistical process, emphasizing the "maturity" of each observation, rather than the vintage of the time series. These models permit forecasts of data that are to be released, as well as "backcasts" of data already published. The methodology may be applied to individual observations, as well as various trend estimators, such as those considered by Orphanides and van Norden (2005).[3] Recent work includes Jacobs and van Norden (2006); Cunningham et al. (2007); Aruoba (2008); Aruoba, Diebold, and Scotti (2008); Garratt, Koop, and Vahey (2008); and Garratt et al. (2008).

This recent literature traces its beginning to Jacobs and van Norden (2006). They argue that previous state-space models built on a transition process for the vintage data plus a set of forecasting equations do not allow adequately rich dynamics in the data-revision process:

> Our formulation of the state-space model is novel in that it defines the measured series as a set of various vintage estimates for a given point in time, rather than a set of estimates from the same vintage. We find this leads to a more parsimonious state-space representation and a cleaner distinction between various aspects of measurement error. It also allows us to augment the model of published data with forecasts in a straightforward way. (p. 3)

In this spirit, we note the differences between using state-space models as estimators of unobserved components such as trends (perhaps across various vintages of real-time data) and as estimators of "true" underlying data. In the former, each datum within a time series of a particular vintage is implicitly assumed to be equally accurately measured; the trend (usually, a time-varying direction vector) is extracted without explicit concern for measurement error, except so far as the robustness of the extracted trend may be explored across vintages. In the latter, each datum within a time series is assumed to have an amount of measurement error that is inversely correlated with the maturity of the datum. Interpreted loosely, the information content of an older datum for a given activity date is asserted to contain more information about the "true" value of that datum than a recent datum for the same activity date.

This modeling framework has been applied by staffs at the Bank of England and the European Central Bank (Cunningham et al., 2007). Their model differs somewhat from that of Jacobs and van Norden and focuses more attention on modeling the measurement-error process, including potential bias and heteroskedasticity, but the underlying philosophy is similar. Our research applies the Jacobs and van Norden framework to U.S. data on quarterly GDP from the Federal Reserve Bank of St. Louis real-time ArchivaL Federal Reserve Economic Data (ALFRED) database.

STATISTICAL FRAMEWORK

The rich modeling framework proposed by Cunningham et al. (2007) allows serial correlation in measurement errors, nonzero correlation between the state of the economy and measurement errors, and maturity-dependent heteroskedasticity in measurement errors. As a consequence of the richness of the statistical specification and the number of dimensions to the data, the estimation is divided into two parts. First, all available data vintages are used to estimate selected parameters governing measurement error bias and variance. Second, the most recently published release is used to estimate the state-space model.

The modeling setup is as follows.[4] Let the data-generating process for the true (unobserved) variable of interest, $y_t, t=1,\ldots,T$, be a simple autoregressive (AR(q)) process:

$$A(L)y_t = \varepsilon_t, \tag{1}$$

where the polynomial is defined in the usual manner and the stationary disturbance is spherical

[3] These trend estimators are discussed in Orphanides and van Norden (2005, Appendix A).

[4] The model follows Jacobs and van Norden (2006) and Cunningham et al. (2007).

UNDERSTANDING REAL-TIME DATA

The table presents a stylized real-time dataset. The columns denote the release date, or vintage, of the data. The rows denote the activity date, or observation date, of the data. Economic data are normally released with a one-period lag; that is, data for January are reported in February. Therefore, the release date, v, lags the activity date, t, by one period.

Each element in the dataset is reported with a subscript identifying the activity date and a (**bold**) superscript identifying the maturity, j. Data of constant maturity are reported along each diagonal.

Stylized Real-Time Dataset

		Vintage (v)						
	Period	$v = 2$	$v = 3$	$v = 4$	…	$v = T-1$	$v = T$	$v = T+1$
Activity date (t)	$t = 1$	$y_1^{\mathbf{1}}$	$y_1^{\mathbf{2}}$	$y_1^{\mathbf{3}}$	…	$y_1^{\mathbf{T-2}}$	$y_1^{\mathbf{T-1}}$	$y_1^{\mathbf{T}}$
	$t = 2$		$y_2^{\mathbf{1}}$	$y_2^{\mathbf{2}}$	…	$y_2^{\mathbf{T-3}}$	$y_2^{\mathbf{T-2}}$	$y_2^{\mathbf{T-1}}$
	$t = 3$			$y_3^{\mathbf{1}}$	…	$y_3^{\mathbf{T-4}}$	$y_3^{\mathbf{T-3}}$	$y_3^{\mathbf{T-2}}$
	⋮				⋱	⋮	⋮	⋮
	$t = T-2$					$y_{T-2}^{\mathbf{1}}$	$y_{T-2}^{\mathbf{2}}$	$y_{T-2}^{\mathbf{3}}$
	$t = T-1$						$y_{T-1}^{\mathbf{1}}$	$y_{T-1}^{\mathbf{2}}$
	$t = T$							$y_T^{\mathbf{1}}$

(homoskedastic), $E(\varepsilon_t) = 0$, $V(\varepsilon_t) = E(\varepsilon_t \varepsilon_t') = \sigma_\varepsilon^2 I$. Trends (deterministic or stochastic) and structural breaks, including regime shifts, are explicitly ruled out (and perhaps have been handled by prefiltering the data).

Measurement-Error Model

Let the data published by the statistical agency be denoted

$$y_t^j,\ t = 1,\ldots,T;\ j = 1,\ldots,J,$$

where t is an activity date and j is the maturity of the data for that activity date. (See the boxed insert.) We assume that initial publication of data for t occurs in period $t+1$, so that $j \geq 1$. Period T is the final revision date for data published in period $T+1$. We assume the published data are decomposable as

$$y_t^j \equiv y_t + c^j + v_t^j, \tag{2}$$

where y_t denotes the true "unobserved" value, c^j denotes a bias in published data of vintage j, and v_t^j is a measurement error.

Previous studies have suggested that data releases tend to be biased estimates of the later releases. Let c^j denote the bias of data at maturity j, such that c^1 is the bias for initially published data. We assume the bias is independent of vintage and solely a function of maturity, j, and that the bias decays according to the rule

$$c^j = c^1(1+\lambda)^{j-1},\ -1 \leq \lambda \leq 0. \tag{3}$$

We assume the measurement error, v_t^j, follows a simple AR(q) process:

$$B(L)v_t^j = \eta_t^j, \tag{4}$$

where $E(\eta_t^j) = 0$. The measurement-error variance is assumed heteroskedastic in maturity and decays toward zero:

$$V\left(\eta_t^j\right) = E\left(\eta_t^j\left(\eta_t^j\right)'\right) = \Sigma_\eta^j,$$

with main diagonal

(5) $$\sigma_{\eta^j}^2 = \sigma_{\eta^1}^2\left(1+\delta\right)^{j-1},\ -1 \le \delta \le 0.$$

It is necessary to designate a maturity at which data are assumed "fully mature." Here, we denote the horizon as N and refer to it below as the "revision horizon." For RGDP data, we set $N = 20$ (5 years of quarterly data). This choice, to some extent, is arbitrary, and hence it is useful to examine the robustness of results to the value chosen. Our choices are guided by visual examination of the revised time series and discussed further in a later section.

STATE-SPACE MODEL

The measurement equation of the state-space model has as its dependent variable a vector of the most recent release of data,

$$y_t^j = \left[y_1^T, y_2^{T-1}, \ldots, y_{T-1}^2, y_T^1\right]^T.$$

The superscript j denotes the vintage of data, measuring y_t on activity date t, which is available at vintage T+1. Note that the maturities of the elements of

$$\left[y_1^T, y_2^{T-1}, \ldots, y_{T-1}^2, y_T^1\right]^T$$

differ—some elements may be the 10th or 20th release of data for a specific activity date, while the last element is the initial release of data for activity period T. The measurement equation equates this vector to the sum of a vector of maturity-related measurement biases, c^j; the unknown true value, y_t; and a measurement error, v_t^T:

(6) $$y_t^j = c^j + \begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix} y_t \\ v_t^T \end{bmatrix} + 0.$$

The transition equation for the state vector is

(7) $$\begin{bmatrix} y_t \\ v_t^T \end{bmatrix} = \begin{bmatrix} \mu \\ 0 \end{bmatrix} + \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}\begin{bmatrix} y_{t-1} \\ v_{t-1}^T \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \eta_t^T \end{bmatrix}$$

with disturbance covariance matrix

(8) $$\begin{bmatrix} \sigma_{\varepsilon^T}^2 & \sigma_{\varepsilon,\eta^{(j,t)}}^2 \\ \sigma_{\varepsilon,\eta^{(j,t)}}^2 & \sigma_{\eta^{(j,t)}}^2 \end{bmatrix}.$$

Note that the variance of η_t^T, denoted $\sigma_{\eta^j}^2$, is a function of t through its dependence on j, the maturity of each datum, reflecting the assumed heteroskedasticity in the measurement-error process.[5] Similarly, the covariance between the shocks to the variable of interest and the measurement error, $\sigma_{\varepsilon\eta^{(j,t)}}^2 = \rho_{\varepsilon,\eta}\sigma_\varepsilon\sigma_{\eta^j}$, is a function of maturity, j.

Cunningham et al. (2007) make the interesting suggestion that the measurement equation may be augmented with auxiliary variables y_t^Z:

(9) $$\begin{pmatrix} y_t^T \\ y_t^Z \end{pmatrix} = \begin{pmatrix} c^j \\ c^Z \end{pmatrix} + \begin{pmatrix} I & 0 & I & 0 \\ \Lambda^Z & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} y_t \\ \vdots \\ y_{t-q+1} \\ v_t \\ \vdots \\ v_{t-p+1} \end{pmatrix} + \begin{bmatrix} 0 \\ v_t^Z \end{bmatrix}.$$

Candidate variables include surveys and/or private-sector measures/forecasts, asserting that private-sector agents already have solved their own variants of the signal-extraction problem. At this time, our model omits the use of auxiliary data.

The estimation is partitioned into two parts. Assuming the measurement equation has been augmented with an auxiliary variable and allowing for AR(1) processes in the transition equation, the parameters to be estimated in the state-space model are

$$\Phi_1 = \left(\alpha_1, \sigma_{v^Z}^2, \mu, \sigma_\varepsilon^2, c^Z, \Lambda^Z\right),$$

conditional on estimated parameters for the measurement error's data-generating process,

$$\Phi_2 = \left(\sigma_{\eta^1}^2, \delta, \beta_1, \rho_{\varepsilon,\eta}, c^1, \lambda\right).$$

The estimation of Φ_2 proceeds assuming that successive revisions to each datum are well-

[5] State-space models with deterministically time-dependent variances are discussed by Durbin and Koopman (2001, pp. 172-74) and Kim and Nelson (1999).

behaved, in the statistical sense that the revisions may be used for estimation.[6] Let W denote a matrix with J rows, in which each row is regarded as a vector of revisions to data of maturity j. The number of columns is $T-N$, that is, the number of published data vectors minus the revision horizon. A general expression for a representative row in the revision matrix W is

$$(10)\quad W(j,.) = y_t^{j+N} - y_t^j,\ j+N<T,\ 1 \le t \le T.$$

Consider $j = 1$ and $N = 20$. In this case, the numbers in the first row of W are

$$W(1,.) = y_t^{1+20} - y_t^1,\ 1+N<T.$$

Similarly, consider $j = 12$ and $N = 20$:

$$W(12,.) = y_t^{12+20} - y_t^{12},\ 12+N<T.$$

Clearly, W has J rows and $T-N$ columns.

Consider an estimator for the bias process,

$$(11)\quad c^j = c^1(1+\lambda)^{j-1},\ -1 \le \lambda \le 0,\ 1 \le j \le J.$$

The row means of W provide sample measures of c^j. The parameters c^1, the mean revision of the initial release, and λ, the revision decay rate, are estimated via generalized method of moments (GMM) subject to the constraint $-1 \le \lambda \le 0$.

Next, we need an estimator for $\rho_{\varepsilon,\eta}$ as part of $\sigma^2_{\varepsilon\eta} = \rho_{\varepsilon,\eta}\sigma_{\varepsilon_i}\sigma_{\eta_i^T}$. Cunningham et al. (2007) propose an estimator based on an approximation to $\rho_{\varepsilon,\eta}$, designated $\rho^*_{y,v}$, calculated as the mean (across the J maturities) of the J correlations between the jth rows of W and the corresponding vector of published data at maturity $j+N$; that is, by the construction of W, $N+j<T-t$.

Finally, estimators are required for $\sigma^2_{\eta^1}$, δ, and β_1 (assuming an AR(1) process in v). A sample estimate of the variance-covariance is obtained as $J^{-1}WW'$. The analytical covariance matrix for the first-order case is

$$(12)\quad V = \left(\frac{\sigma^2_{\eta^1}}{1-(1+\delta)\beta_1^2}\right)\times \begin{pmatrix} 1 & (1+\delta)\beta_1 & \cdots & (1+\delta)^{J-1}\beta_1^{J-1} \\ (1+\delta)\beta_1 & (1+\delta) & \cdots & (1+\delta)^{J-1}\beta_1^{J-2} \\ \vdots & \vdots & \ddots & \vdots \\ (1+\delta)^{J-1}\beta_1^{J-1} & (1+\delta)^{J-1}\beta_1^{J-2} & \cdots & (1+\delta)^{J-1} \end{pmatrix},$$

and we estimate $\sigma^2_{\eta^1}$, δ, and β_1 via GMM by minimizing

$$(13)\quad \left(vecV - vec\hat{V}\right)'\left(vecV - vec\hat{V}\right).$$

Cunningham et al. (2007) suggest methods to obtain covariance matrices for higher lag orders.

SIMULATION RESULTS

We conduct Monte Carlo simulations to explore the ability of the state-space framework to extract a "true" series from a "published" series that has been contaminated with measurement error.

The simulations evaluate the ability of the model's state-space vector $[\hat{y}_t, \hat{v}_t]$ to track the vector of true values, y_t, relative to the tracking ability of the vector of most recently "published" values, y_t^T. For each parameterization, $T = 100$ and we calculate 1,000 replications.

The specification of the experiment is as follows:

- The "true" data:

$$y_t = \alpha y_{t-1} + \varepsilon_t,\quad t = 2,\ldots,T,\ \varepsilon_t \sim N(0,1),\ y_1 = \varepsilon_1$$
$$(\text{that is, } y_0 = 0).$$

- The "published" data:

$$y_t^j = y_t + v_t^j,\quad t = 1,\ldots,T,$$

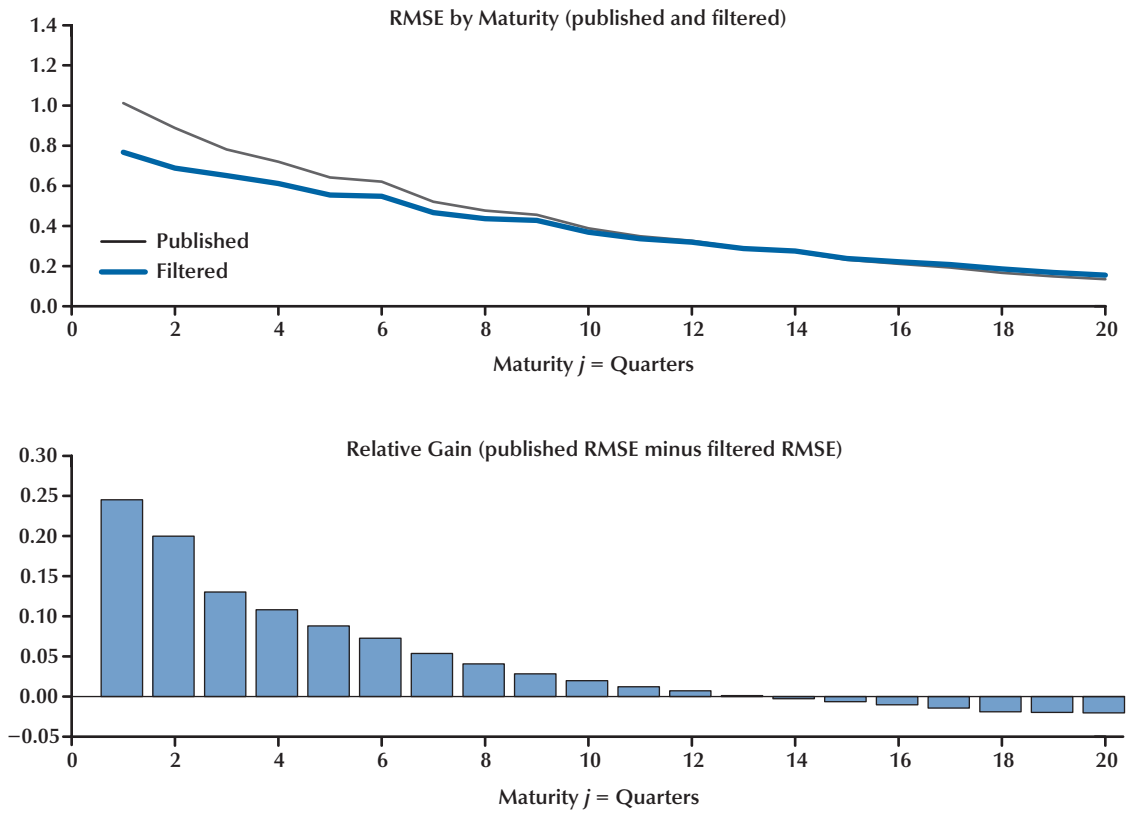
  where the superscripts t and j denote, respectively, the activity date and maturity of the most recently published data.

- The measurement error:

[6] Hereafter, this exercise is conditional on the revision horizon N.

Figure 1

RMSE by Maturity, *j*, Model-Simulation Parameters: α = 0.60, β = 0.10, $\rho_{\varepsilon\eta}$ = –0.50



$$v_t^j = \beta v_{t-1}^j + \eta_t^j, \quad t = 2,\ldots,T; \quad \eta_t^j \sim N\left(0,\sigma_{\eta_t^j}^2\right);$$

$$\sigma_{\eta^1}^2 = 1; \quad \sigma_{\eta_t^j}^2 = \sigma_{\eta^1}^2 \left(1+\delta\right)^{j-1},$$

$v_1^j = \eta_1^j$ (that is, $v_0^j = 0$) and with $\delta = -0.06$.

- The covariance between the state of the economy and the measurement error:

$$\operatorname{cov}\left(\varepsilon_t, \eta_t^j\right) = \rho_{\varepsilon,\eta}\sigma_\varepsilon\sigma_{\eta_t^j}.$$

Figure 1 shows root mean square forecast errors (RMSEs) of the "published" values, $y_t - y_t^T$, and the filtered values, $y_t - \hat{y}_t$, for one parameterization. The top panel shows the RMSE at each maturity; the bottom panel shows the difference between the filtered RMSEs and published RMSEs. The figure indicates that the filtered values are better estimates of the true values for the first 13 maturities, after which time the filtered values cease to provide an advantage.

Table 1 provides corresponding results.[7] The first three columns report varying parameterizations. The fourth and fifth columns report the improvement due to the state-space filter for data maturities 1 and 10, respectively. At maturity 1, the RMSEs of filtered estimates are approximately

[7] In some respects, the results presented in Table 1 are similar to those in Cunningham et al. (2007), while others are puzzlingly different. The comparable table in Cunningham et al. (2007, Table C) displays a sharp deterioration of the model's advantage as the covariance shock to the economy and measurement error decreases. Additionally, the gains reported are much larger and more persistent; in some cases, the filtered RMSEs are close to 50 percent of the published RMSEs at maturity 1 and remain so at maturity 9. The gains from filtering last at least 18 but sometimes over 100 periods. At this time we have not resolved these discrepancies; but we thank the authors for graciously providing their simulation code.

Table 1

Measurement Accuracy Improvement Due to the State-Space Filter

Parameterization			Improvement due to state-space filter		
			$RMSE_{filtered}/RMSE_{published}$		Earliest maturity at which
$\rho_{\varepsilon,\eta}$	α	β	At maturity 1	At maturity 10	$RMSE_{published} < RMSE_{filtered}$
0.5	0.1	0.1	0.7179	1.1338	8
0.5	0.1	0.6	0.6489	1.0762	9
0.5	0.6	0.1	0.7500	1.1437	8
0.5	0.6	0.6	0.7344	1.1581	7
0	0.1	0.1	0.7179	1.0357	10
0	0.1	0.6	0.6479	1.0277	9
0	0.6	0.1	0.7493	1.0226	10
0	0.6	0.6	0.7337	1.0478	9
–0.5	0.1	0.1	0.7177	0.9378	15
–0.5	0.1	0.6	0.6537	0.9377	20
–0.5	0.6	0.1	0.7578	0.9494	14
–0.5	0.6	0.6	0.7331	0.9431	14

70 percent of the RMSEs obtained when using the most recently "published" data. The values range from 64 to 75 percent improvement depending on the parameterization. The sixth column reports the earliest maturity at which the filtered values cease to provide an advantage; these values range from 8 to 20 periods, with an average of 11 periods.

Our simulations suggest that the state-space framework may promise significant gains in measurement accuracy for recently released data if actual data are well behaved and tend to follow a low-order AR process. Previous studies suggest this might be reasonable for RGDP.

EMPIRICAL MODEL

Our empirical work examines vintage data of the annualized growth rate of quarterly RGDP constructed with data from the Federal Reserve Bank of St. Louis ALFRED database, specifically, nominal GDP and the implicit price deflator (GDPDEF) data.[8] The construction of RGDP accounts for changes in the base year of GDPDEF, as to maintain the correct interactions between the base year and subsequent vintages. Thus, in this vintage RGDP matrix, the most recently published data vector matches the data available from the Bureau of Economic Analysis (BEA). The specifics of the process are described in Appendix B.

Estimation proceeds in two steps: First, we estimate the parameters of the measurement-error process,

$$\Phi_2 = \left(\sigma^2_{\eta^1}, \delta, \beta_1, \rho_{\varepsilon,\eta}, c^1, \lambda\right).$$

Second, conditional on these parameters, we estimate the parameters of the state-space model (omitting any auxiliary data),

$$\Phi_1 = \left(\alpha_1, \sigma^2_{v^Z}, \mu, \sigma^2_{\varepsilon}\right).$$

Cunningham et al. (2007) note one reason to use this two-step procedure is that identification conditions may fail if all parameters were estimated together.[9] Moreover, the framework set

[8] The adoption of a chain-weighted price index in the middle of the sample adds an additional dynamic to the RGDP revision process. It would be ideal to use only post chain-weighted data; however the sample size is not sufficient for estimation.

[9] Cunningham et al. (2007) do not explore the satisfaction and/or violation of the relevant conditions; neither have we, although so doing seems a worthwhile task, to say the least.

Table 2

Estimated Revision-Bias Parameters

RGDP	Estimate	Lower bound	Upper bound
c^1	0.5793	0.3317	0.8269
λ	–0.2828	–0.4515	–0.1141

NOTE: Upper and lower bounds represent 95 percent confidence intervals (CIs).

Table 3

Estimated $\hat{v}$ Parameters

RGDP	Estimate	Lower bound	Upper bound
Initial variance, $\sigma^2_{\eta^1}$	2.7002	2.3988	3.0017
Variance decay, δ	–0.0786	–0.0931	–0.0641
First-order serial correlation, β_1	0.2004	0.1451	0.2557
Correlation with mature data, ρ^*_{yv}	0.3181	0.1399	0.4963

NOTE: Upper and lower bounds represent 95 percent CIs.

forth requires only the most recently published vintage of data, joint estimation of the parameters would require inputting the entire history of revisions into the model.

Estimation of Φ_2 Parameters

As noted previously, the first step of the estimation is to choose the revision horizon. Here it is $N = 20$ (5 years). Our choice is explored in Appendix A. We input the W matrix (produced by equation (10)) into equation (11) to estimate values for the mean revision of the initial release, c^1, and the revision decay parameter, λ. For robustness purposes, Figure A2 shows the estimated and actual values of c^j at different horizons.

Table 2 reports the parameters estimated via GMM. The mean revision to the initial release is statistically different from zero. The initial release to RGDP is, on average, 0.57 percentage points lower than RGDP reported five years later. The revision decay parameter describes the rate at which revisions decay as the data mature: At revision maturity 2, RGDP is estimated to be 28 percent lower than the initial revision.

The next step is to calculate the correlation between the measurement error and the "true" unobserved state of the economy, $\rho_{\varepsilon,\eta}$. Because we do not observe the measurement error or the "true" state, we continue to use revisions to data of maturity j as a proxy of the measurement error. We assume data reported with maturity $j+N$ are good estimates of the "true" state of the economy. The estimated mean correlation between the revisions of maturity j and the reported values at $j+N$ are used as an estimate for $\rho_{\varepsilon,\eta}$ denoted $\rho^*_{y,v}$; that is,

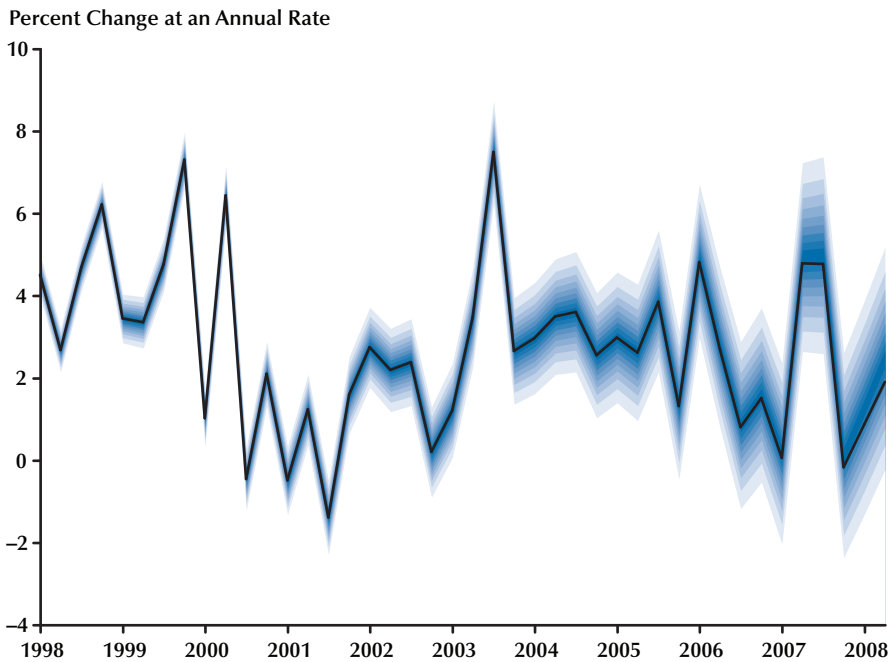
$$\rho_{\varepsilon,\eta} \approx \rho^*_{y,v} = \frac{1}{J}\sum_{j=1}^{J}\rho_{y^{j+N},W(j,.)}.$$

The last row of Table 3 reports estimates for ρ^*_{yv}. The correlation between revisions to the data and the estimated "true" state are positive; although it is not reported, the correlation is positive for all j. Appendix Figure A3 explores the choice of the revision horizon: The values of ρ^*_{yv} stabilize for sufficiently large revision horizons.

The final set of first-stage estimates—the serial correlation between revisions, the initial

Figure 2

Actual and Filtered RGDP Growth (vintage 07/31/2008)



NOTE: The fan chart depicts the probability of various outcomes for RGDP after the data are "fully revised" (i.e., data reported in 5 years). Fully revised data are expected to fall within the fan chart 90 percent of the time. Each pair of shaded regions indicates an additional 10 percent CI.

variance of revisions, and the variance decay rate, $\beta_1, \sigma^2_{\varepsilon_1}, \delta$, respectively—are derived from the variance-covariance matrix of W, denoted $\hat{V}$.

The parameters are estimated per equations (12) and (13). Table 3 reports the estimated parameters for our preferred N. All estimates are significantly different from zero. Notice that the first order–serial correlation in the revisions is positive. The initial variance is markedly higher from that assumed in the simulation. However, the variance decay parameter is –0.07, which is close to the –0.06.value used in the simulation.

Estimation of Φ_1 Parameters

Using the parameters estimated in the previous section, the vector of recently published data is put into the state-space model.[10] The parameter driving the state-space model's covariance matrix, σ^2_ε, the variance of the shock to the AR(q) data-generating process for the "true" data, is estimated to be 3.55 (0.55); for U.K. investment data, Cunningham et al. (2007) report an error variance of 3.22 (0.67). Estimation results are shown in Figure 2. The solid black line is the most recently published data, the darkest band is the mean filtered value, and the outermost band is the 90 percent confidence interval (CI). As the variance of the revisions decay, so does the CI.

The RGDP growth rate at the most recent data point, 2008:Q2, was initially published as 1.89 percent on July 31, 2008. The estimated value is

[10] Estimation of the model is problematic. Although the data-generating processes for both the "true" data and the measurement error are initially asserted to be AR(q), in the model the AR(q) parameters are not identified. The parameters are also omitted from estimation in Cunningham et al. (2007). Absent promising findings in the next section, far more estimation is necessary before confidence may be placed in such results.

2.47 percent, with 90 percent CIs between 5.17 percent and 10.23 percent. As of the February 27, 2009, release, 2008:Q2 RGDP was reported as 2.83 percent.

This state-space modeling framework shares features with multivariate stochastic volatility models. Harvey, Ruiz, and Shephard (1994) introduce such a model as an alternative to generalized autoregressive conditional heteroskedasticity (GARCH) models for high-frequency data. However, their model's problem differs from the current model. In their model, "multivariate" refers to four countries' exchange rates, modeled together, rather than 20 or more maturities of a single activity date. Their problem is similar to the current model, though, to the extent that the stochastic variance is assumed to follow an AR(1) process. This line of econometrics deserves further investigation.

REAL-TIME MODEL EVALUATION

Using real-time RGDP data series to evaluate model accuracy follows closely the methodology of our simulation exercise. The main restriction with the actual data is that we do not observe the true values of each datum. We proxy the true values from data that have become "fully mature" at time $t+N$ (where $N = 20$). Our real-time sample is restricted to those vintages of data with 10 years of data preceding (to estimate the parameters in Φ_2) and 5 years of data following (to evaluate the forecast) the vintages of interest. This exercise uses the data range 1985:Q4–2003:Q2 and vintages between v_0 = 01/30/2002 and v_k = 07/31/2003 as the most recently published data. This does not limit our ability to make real-time forecasts with the current data; however, it does inhibit us from testing the forecasting performance for 5 years.

We estimate the model for vintages $v_0, v_1, \ldots, v_k$ independently, keeping the number of observations, and maturities, fixed across vintages. For each successive release, we omit the oldest datum and add the most recent. This process corresponds nicely with the idea of running k iterations of the model simulation. The metric used to evaluate the model performance is the ratio of the RMSE using the filtered datum as a predictor of the "mature" datum relative to the RMSE using the published datum as a predictor of the "mature" datum.

As described previously, the first stage in the real-time forecasting exercise is to estimate the parameters in Φ_2. Given a revision horizon of $N = 20$ (5 years) for RGDP, at least 10 years of vintage data are used to estimate the W matrix and corresponding parameters.[11] For each successive vintage, we update the dataset (i.e., add a column to the W matrix) and reestimate the parameters in Φ_2.

The stability of each parameter is assessed in Figure 3. The horizontal axis reports the vintage. 01/30/02 indicates that data available only on or before January 30, 2002, were used in the estimation of the parameters, thus the final column of the revision matrix W contains revisions to data released 20 quarters prior. The latest data points are identical to those reported in Tables 2 and 3.

The mean revision of the initial release (top-left panel of Figure 3) steadily decreases as the real-time sample includes more recent data, indicating some improvement in the BEA's ability to report less-biased estimates of the "true" values. Conversely, the initial variance of the measurement error (left-middle panel) increases, indicating increased uncertainty around the initial estimates. The decay parameters corresponding to the initial mean and variance of the revisions are reported by λ and δ, respectively.

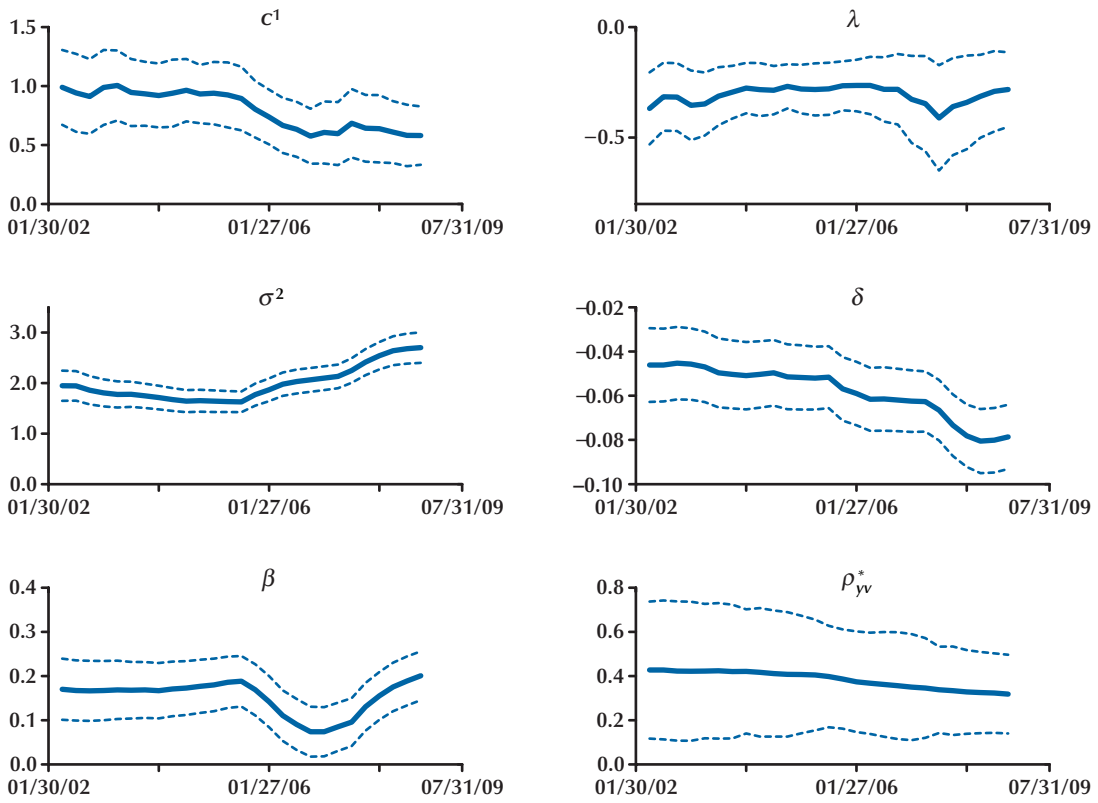
The parameters in Figure 3 are displayed for all available vintages; however, the real-time forecasting exercise uses only the vintages for which the fully mature data are available. As noted earlier, this reduces the real-time sample to only the first seven vintages of data. For example, the mature values for January 30, 2002, data are reported after 20 quarters, on January 31, 2007.

The top panel of Figure 4 plots the RMSEs of the published data and of the filtered values. The bars in the bottom panel measure the difference

[11] We require that the W matrix have at least as many vintages (or columns) as it has maturities (or rows). The calculation in equation (10) requires at least N vintages of data as well as the observed "true" values. In other words, for every vintage, v, we must also observe the data at $v+N$.

Figure 3

Estimated Φ_2 Parameters in Real Time



between the two series.[12] In some ways, the results are similar to those simulated in Figure 1: The filtered values tend to be superior estimates for the first 6 quarters and then diminish. Unlike in the simulation, however, the transition between quarters is not particularly smooth. In the top panel of Figure 4, the RMSEs of both series actually increase for data maturities 3 through 6 and steeply decline thereafter. According to the bars in the bottom panel of Figure 3, for maturity 2, the data from the filtered values show greater improvement than those from the initial release.

For further examination, Table 4 reports the improvement due to the state-space filter for the seven vintages. The first two columns report the vintages of the published data and the fully mature data, respectively. The remaining five columns report the improvement due to the state-space filter for data of varying maturities. The bottom row of the table reports the average across the seven vintages. During the first year, the average RMSE of the filtered values is 87 percent of the average RMSE of the published data. The filtered values most improve the data published July 31, 2003: The RMSE of the filtered values is 48 percent of the RMSE of the published data.[13] For data 2 years old, there is only modest improvement: The RMSE of the filtered values is 97 percent of

[12] The following results should be interpreted with some caution; they are constrained by only seven consecutive releases of data and additional data points may drastically alter the results.

[13] This outlier is driven by a particularly inaccurate initial release of 2.37 percent for 2003:Q1, the filtered value was 3.74 percent, and the value on January 31, 2007, was 3.46 percent.

Figure 4

Real-Time Model Performance, RGDP

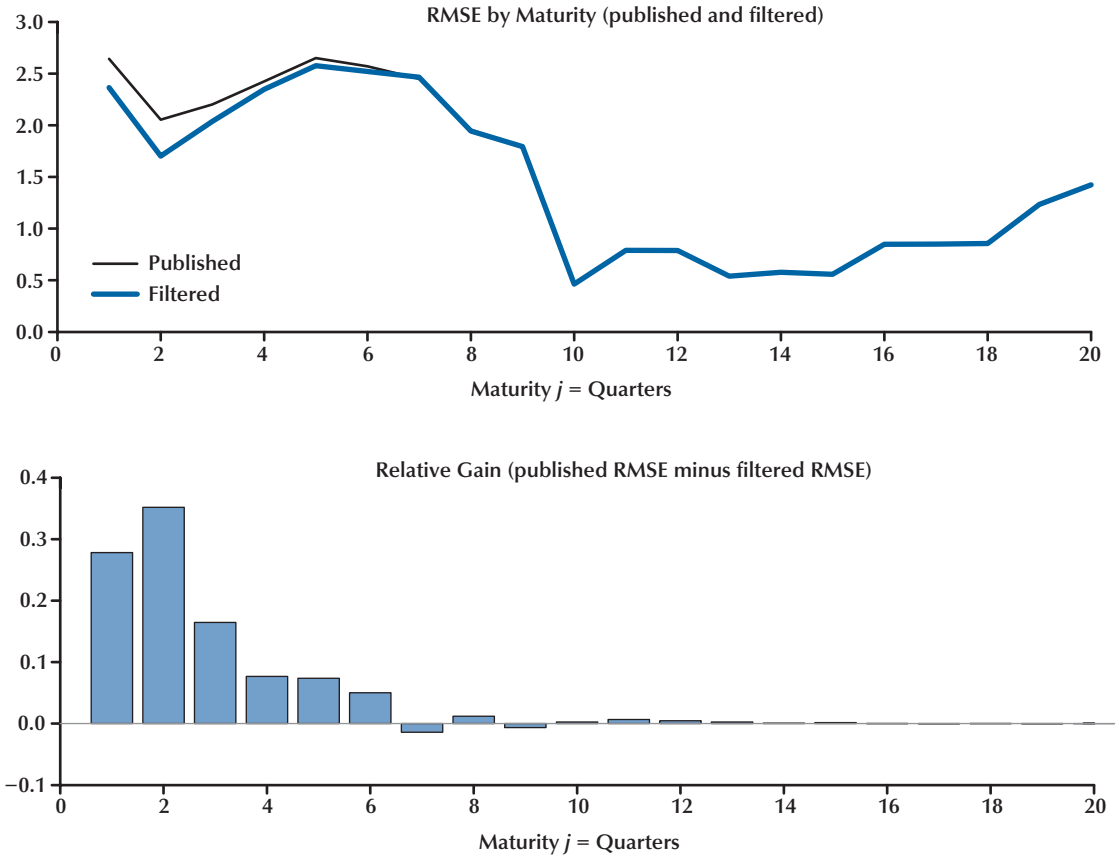


Table 4

Real-Time Model Performance

		Improvement due to state-space filter ($RMSE_{filtered}/RMSE_{published}$)				
Release of published data	Release of fully mature data	Maturities 1-4 (Year 1)	Maturities 5-8 (Year 2)	Maturities 9-12 (Year 3)	Maturities 13-16 (Year 4)	Maturities 17-20 (Year 5)
1/30/2002	1/31/2007	0.9387	0.8561	1.0018	1.0018	1.0002
4/26/2002	4/27/2007	0.8933	0.9988	0.9885	0.9957	1.0002
7/31/2002	7/27/2007	0.9056	0.9601	1.0179	0.9942	1.0003
10/31/2002	10/31/2007	0.9184	0.9980	1.0128	0.9977	1.0000
1/30/2003	1/30/2008	0.9809	0.9725	1.0083	0.9985	1.0000
4/25/2003	4/30/2008	0.9900	0.9815	1.0254	0.9979	0.9988
7/31/2003	7/31/2008	0.4852	1.0039	1.0032	0.9997	0.9996
Average		0.8731	0.9673	1.0083	0.9979	0.9999

the RMSE of the published data. Thus, after data have been revised for 2 years, there is no apparent gain from the filtered values.

In addition to the improved estimates of the true values, as measured by the RMSEs, the filtered values also provide CIs, which are not provided by data releases. The CIs indicate the extent to which incoming data are likely to be revised, providing some assessment as to how much weight to assign to each datum. Of the 504 mature data points observed, approximately 83 percent fell within the 90 percent CI and approximately 41 percent within the 50 percent CI. These numbers seem reasonable, as the sample consists of seven consecutive vintages—meaning any given outlier could be repeated up to seven times within our sample.

CONCLUSION

A long line of papers has explored methods to pool vintages of economic data, seeking to extract the true (or, at least strongest) underlying signal for a variable of interest. A recent, and likely fruitful, path is to introduce a cohort-style analysis that examines the revisions as a function of the age of the data and estimates the "true" unobservable values via a state-space framework. Here, we have begun the application of such techniques to U.S. data, specifically using a measure of RGDP. The framework is equally applicable to quarterly or monthly data, although we have not yet considered the case of mixed frequencies (including when monthly observations are published quarterly, such as for GDP).

Monte Carlo experiments suggest that, for a wide range of parameter values in AR data-generating processes, the framework explored here may be able to extract estimates of recent values of economic variables and reduce uncertainty by as much as 30 percent. Obviously, empirical application of such techniques introduces statistical challenges when pooling data across cohorts. The "revision horizon," to a large extent, is an arbitrary selection, and robustness experiments are required. Further, if underlying unobserved true data are to be recovered as a state vector, issues regarding the lack of statistical identification require further exploration. It appears, however, even with these caveats, the modeling framework does provide estimators for two important variances—the variance of the empirical measurement error embedded in each published datum and the variance of the data-generating process of the true underlying economic variable.

Real-time experiments, albeit with limited data, suggest that uncertainty in RGDP estimates appear to be reduced by close to 10 percent at early maturities. In addition, CIs extracted from the model provide information unattainable from data releases alone. Both, perhaps, will assist economists and policymakers, by providing a set of "revision CIs" around releases of incoming data. One limitation of the application of the methodology is the large amount of data required to produce and evaluate estimates in real time. Nonfarm payroll employment data, with monthly revisions and a long release history, is a good candidate for the application of this methodology.

REFERENCES

Aruoba, S. Borağan. "Data Revisions Are Not Well-Behaved." *Journal of Money, Credit and Banking*, March-April 2008, *40*(2-3), pp. 319-40.

Aruoba, S. Borağan; Diebold, Francis X. and Scotti, Chiara. "Real-Time Measurement of Business Conditions." Unpublished manuscript, April 2007; 2008 version: Working Paper 08-19, Federal Reserve Bank of Philadelphia; www.philadelphiafed.org/research-and-data/publications/working-papers/2008/wp08-19.pdf.

Boskin, Michael J. "Getting the 21st Century GDP Right: Progress and Challenges." *American Economic Review*, May 2000, *90*(2), *AEA Papers and Proceedings*, pp. 247-52.

Croushore, Dean. "Revisions to PCE Inflation Measures: Implications for Monetary Policy." Unpublished manuscript, University of Richmond, November 2007.

Croushore, Dean and Stark, Tom. "A Funny Thing Happened on the Way to the Data Bank: A Real-Time Data Set for Macroeconomists." Federal Reserve Bank of Philadelphia *Business Review*, September/October 2000, pp. 15-27; www.phil.frb.org/research-and-data/publications/business-review/2000/september-october/brso00dc.pdf.

Croushore, Dean and Stark, Tom. "Data Revisions and the Identification of Monetary Policy Shocks." *Journal of Econometrics*, November 2001, *105*(1), pp. 111-30.

Cunningham, Alastair; Eklund, Jana; Jeffrey, Chris; Kapetanios, George and Labhard, Vincent. "A State Space Approach to Extracting the Signal from Uncertain Data." Working Paper 336, Bank of England, November 2007; www.bankofengland.co.uk/publications/workingpapers/wp336.pdf.

Diebold, Francis X. and Rudebusch, Glenn D. "Forecasting Output with the Composite Leading Index: A Real-Time Analysis." *Journal of the American Statistical Association*, September 1991, *86*(415), pp. 603-10.

Durbin, James and Koopman, Siem J. *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press, 2001.

Faust, Jon; Rogers, John H. and Wright, Jonathan. "News and Noise in G-7 GDP Announcements." *Journal of Money, Credit, and Banking*, June 2008, *37*(3), pp. 403-19.

Garratt, Anthony; Koop, Gary and Vahey, Shaun P. "Forecasting Substantial Data Revisions in the Presence of Model Uncertainty." *Economic Journal*, July 2008, *118*(530), pp. 1128-44.

Garratt, Anthony; Lee, Kevin; Mise, Emi and Shields, Kalvinder. "Real Time Representations of the Output Gap." *Review of Economics and Statistics*, November 2008, *90*(4), pp. 792-804.

Grimm, Bruce T. and Weadock, Teresa L. "Gross Domestic Product: Revisions and Source Data." *Survey of Current Business*, February 2008, *86*(8), pp. 11-15.

Harvey, Andrew; Ruiz, Esther and Shephard, Neil. "Multivariate Stochastic Variance Models." *Review of Economic Studies*, April 1994, *60*(2), pp. 247-64.

Howrey, E. Philip. "The Use of Preliminary Data in Econometric Forecasting." *Review of Economics and Statistics*, May 1978, *60*(2), pp. 193-200.

Howrey, E. Philip. "Data Revision, Reconstruction, and Prediction: An Application to Inventory Investment." *Review of Economics and Statistics*, August 1978, *66*(3), pp. 386-93.

Jacobs, Jan P.A.M. and van Norden, Simon. "Modeling Data Revisions: Measurement Error and Dynamics of 'True' Values." CCSO Working Paper 2006/07, CCSO Centre for Economic Research, December 2006; www.eco.rug.nl/ccso/quarterly/200607.pdf.

Kim, Chang-Jin and Nelson, Charles R. *State-Space Models with Regime Switching*. Cambridge, MA: MIT Press, 1999.

Kishor, N. Kundan and Koenig, Evan F. "VAR Estimation and Forecasting When Data Are Subject to Revision." Working Paper 0501, Federal Reserve Bank of Dallas, February 2005; http://dallasfed.org/research/papers/2005/wp0501.pdf.

Kozicki, Sharon. "How Do Data Revisions Affect the Evaluation and Conduct of Monetary Policy?" Federal Reserve Bank of Kansas City *Economic Review*, First Quarter 2004, pp. 5-37.

Landerfeld, J. Steven; Seskin, Eugune P. and Fraumeni, Barbara M. "Taking the Pulse of the Economy: Measuring GDP." *Journal of Economic Perspectives*, Spring 2008, *22*(2), pp. 193-216.

Mankiw, N. Gregory and Shapiro, Matthew D. "News or Noise? An Analysis of GNP Revisions." NBER Working Paper No. 1939, National Bureau of Economic Research, June 1986; www.nber.org/papers/w1939.pdf?new_window=1.

Orphanides, Athanasios and van Norden, Simon. "The Unreliability of Output-Gap Estimates in Real Time." *Review of Economics and Statistics*, November 2002, *84*(4), pp. 569-83.

Orphanides, Athanasios and van Norden, Simon. "The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time." *Journal of Money, Credit, and Banking*, June 2005, *37*(3), pp. 583-601.

Sargent, Thomas. "Two Models of Measurements and the Investment Accelerator." *Journal of Political Economy*, April 1989, *97*(2), pp. 251-87.

Stock, James H. and Watson, Mark W. "Why Has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit, and Banking*, January 2007, *39*(1), pp. 3-33.

APPENDIX A: SELECTION OF REVISION HORIZON (*N*)

Figure A1 shows the rows of the *W* matrix produced by equation (10) for RGDP over time at two different maturities, *j*, and horizons, *N*. The top panel shows the revision to the initial release after 1 quarter; the second panel shows the revision to the 20th release after 1 quarter. As expected, the revisions to the 20th release after 1 quarter tend to be zero. When the horizon is extended from 1 quarter to 5 years (bottom two panels), the 20th release does exhibit revision. Figure A2 shows the estimated and actual revision process of data subject to a revision horizon $N = 1$, 5, 10, and 20. Figure A3 shows the correlation between revisions and the "true" data subject to a revision horizon.

Figure A1

Revisions to Initial Release and 5-Year-Old RGDP Data at Different Horizons

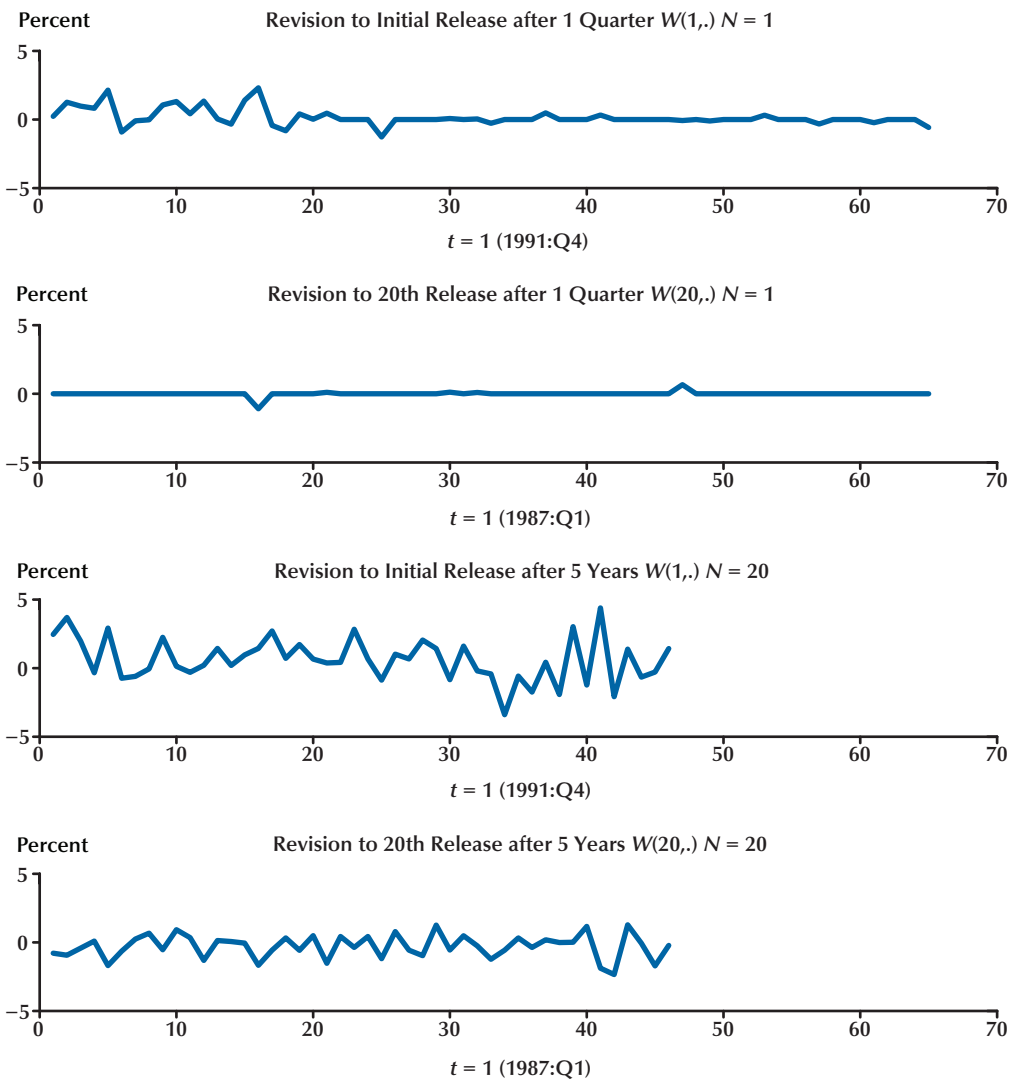


Figure A2

Mean Revisions to RGDP by Maturity, *j*, at Revision Horizon *N*

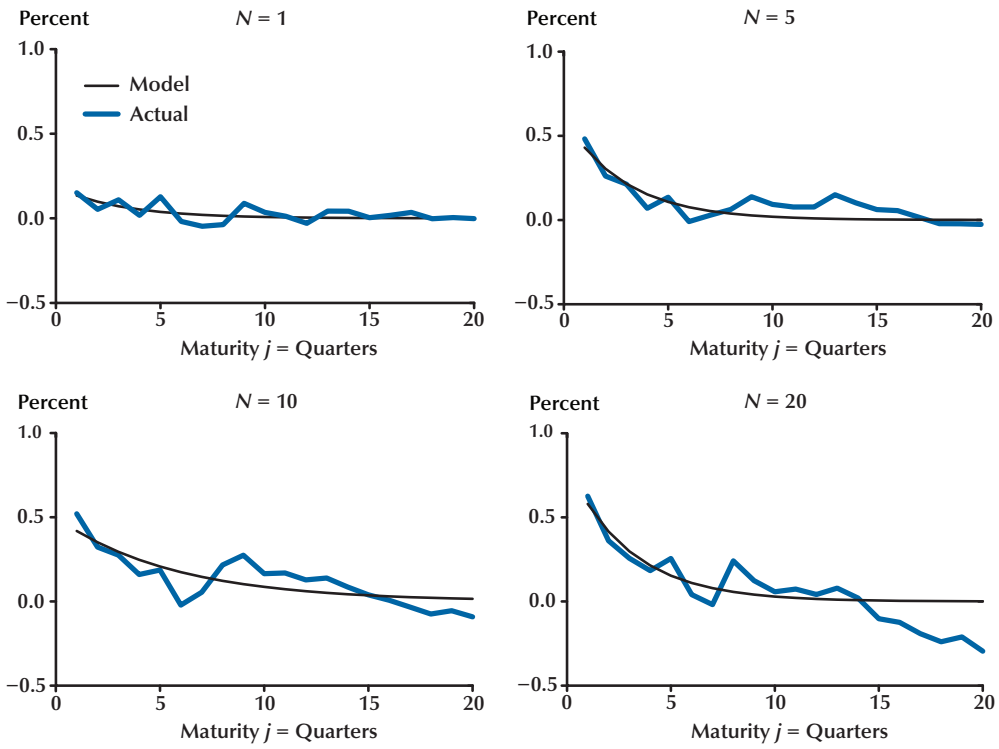
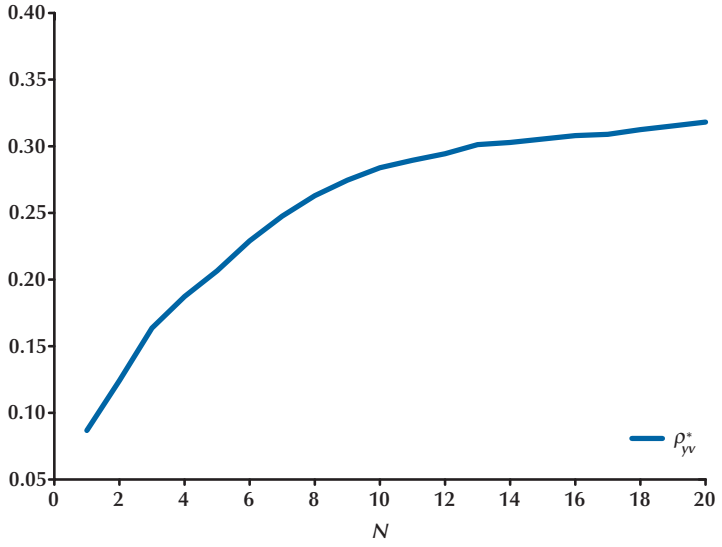


Figure A3

Correlation of Revisions Between Maturity, *j*, and Published Estimates at Maturity, *j*+*N*



APPENDIX B: ABOUT THE DATA

How the Real GDP Data Are Created

The Federal Reserve Bank of St. Louis ALFRED database allows researchers to retrieve vintage versions of economic data that were available on specific dates in history. Most data are available in real and nominal terms. If a researcher is interested in one vintage of data, the real series may be suitable; however, in our case we are interested in all vintages of real GDP. As reported in ALFRED, the unit of measure on the real series changes by vintage. For example, between December 4, 1991, and January 18, 1996, real GDP is reported in billions of 1987 dollars, whereas between January 19, 1996, and October 28, 1999, the series is reported in billions of chained 1992 dollars. Due to changes in the deflator, it is not suitable to obtain the real series from ALFRED and simply calculate the revisions. As an alternative, the nominal GDP (GDPA), GDP, and implicit price deflator (GDPDEF) series are used to create a vintage real GDP series.[14]

As with real GDP, the unit of measure of GDPDEF changes across vintages. Therefore, before deflating GDPA, GDPDEF must be reindexed. The data available in ALFRED are "as reported," meaning the base year varies from 1987 = 100 for vintages before January 18, 1996, to 2000 = 100 for vintages after December 12, 2003. Further complicating the issue, the data released in the base years (1985, 1992, 1996, and 2000) are also subject to revision; therefore the indexing of GDPDEF can also change between vintages within the same base year. Because we are interested in revisions to the data resulting from new information, and not simply changes in the base year, we reindex all GDPDEF data to a constant base year. To match the new series to the most recently reported data, we choose to index all of the data by setting 2000 = 100 in the July 31, 2008, release. We denote the new deflator series DEFL.

The real GDP series are constructed by multiplying each date and vintage of the GDPA by the corresponding date and vintage of DEFL. After deflating the data, annualized growth rates of each vintage are calculated, and we denote the resulting series RGDP.

Because the models are not well suited for mixed-frequency data,[15] we elect to use only the data vintages in which a new advance estimate is released. Consistent with our dataset, the first maturity ($n = 1$) in national income and product accounts (NIPA) data is the advance estimate. In the NIPA data from ALFRED, the preliminary estimate would be the second maturity; however, we omit this vintage, as well as the final estimate. We label the fourth release, which is released at the same time as the subsequent quarter's advance estimate, as the second maturity ($n = 2$).

Table B1 presents a stylized real-time dataset after the preliminary and final vintages have been removed from the data. The columns denote the data vintages; the rows denote the dates of the observations. For descriptive purposes, each element in the dataset is reported with a superscript identifying the maturity, j, of the observation.

The analysis in this paper hinges on the value chosen for the maturity horizon, or "look-ahead distance," denoted J. The value of J is the assumed horizon at which the data are assumed to be true, in that no further revisions to the data will occur. This paper is absent a discussion about the appropriate horizon. Our visual inspection of the data, summarized in Appendix A, and data limitations lead us to set a 5-year horizon ($J = 20$) for GDP and RGDP. For robustness purposes, in Figures A1, A2, and A3 all parameters in Φ_2 are reported for alternative values for J.

[14] The GDPDEF is chosen over the preferred chain-type price index (GDPCTPI) when available. The oldest vintage for GDPDEF is December 4, 1991, whereas the oldest vintage for GDPCTPI is January 19, 1996.

[15] The BEA releases the quarterly GDP series at a monthly frequency: The first release is the advance, the second the preliminary, and the third the final release.

Table B1

Real-Time Dataset: Annualized Growth Rate of Real GDP

	Vintage (v)						
Activity date (t)	1999:Q3	1999:Q4	2000:Q1	…	2008:Q1	2008:Q2	2008:Q3
1999:Q2	5.6^{1}	4.7^{2}	4.7^{3}	…	6.3^{35}	6.3^{36}	6.3^{37}
1999:Q3		6.8^{1}	7.9^{2}	…	7.7^{34}	7.7^{35}	7.7^{36}
1999:Q4			10.1^{1}	…	11.0^{33}	11.0^{34}	11.0^{35}
⋮				⋱	⋮	⋮	⋮
2007:Q4					5.8^{1}	5.5^{2}	4.9^{3}
2008:Q1						5.9^{1}	6.10^{2}
2008:Q2							4.16^{1}

NOTE: Superscripts denote maturity, j. Following the notation in the paper, y_t^j denotes y at time t at maturity, j, or $y_{2007:Q4}^3$ equals 4.9 percent.

Why Are NIPA Data Revised?

Clearly, revisions to NIPA data are not caused by statisticians at the BEA finding computational errors and fixing them. Two main causes of such revisions to NIPA data are that over shorter horizons new data become available (thus prompting revisions) and over longer horizons methodology changes. Statisticians and economists at the BEA are well aware of these problems and over time have made significant updates to the data collection and publication process. At the same time, this paper assumes that by mining the data and revision process we can more accurately predict the true values of a series of interest. We make this assumption not because of any inadequacy of the BEA's work, but rather because of the complexity of the task.

Short-term data revisions are largely a result of the tradeoff faced by the BEA. On one hand, there is pressure for timely releases of information; on the other hand, there is an assumption that the data released accurately measure the underlying variable of interest. Because of the desire for timely estimates, the BEA releases their first, or "advance," estimate with only 75 percent of data for the past quarter (Landefeld et al., 2008). The estimates of the true value are revised as more data become available. Table B2 outlines the four data types used to construct the GDP series as well as the total share of each for 2003:Q3, as reported Grimm and Weadock (2008) in the *Survey of Current Business*. Trend-based data are imputed data; complete data are data that have been reported for the quarter for all three months of the quarter; monthly trend-based data include two months of data and imputed-data for the third month of the quarter; and revised data are simply revised estimates of the complete data. Notice that the advance estimate ($n = 1$) does not contain any revised data and less than half of the data is complete, whereas over three-fourths of the data in the final release ($n \approx 2$) is complete or has been revised. At the time of the annual revision,[16] over 90 percent of the data is complete or has been revised. Detailed information on the data sources, revision process, and methodology used to create the NIPA data are provided by Landefeld et al. (2008).

[16] The maturity of these data is a function of t. For Q1 data the annual revision will occur at $n \approx 4$; for Q2 data at $n \approx 3$; for Q3 data at $n \approx 2$; and Q4 data will not be subject to an annual revision until the next year, or $n \approx 5$.

Table B2

Data Sources for Short-Term Revisions to GDP (percent)

Data source	Share of 2003:Q3 GDP	
	Advance estimate	Final estimate
Trend-based data	25.1	20.9
Monthly and trend-based data	29.7	1.2
Complete data	45.3	8.4
Revised data	—	69.5

SOURCE: Grimm and Weadock (2008).

In addition to problems caused by the lack of data available, challenges exist in regard to quantifying the actions of economic agents, such as the growth in the service sector, identifying new products as they enter the economy, and quality improvements for existing products (see Boskin, 2000). Because of the large scale of these problems, the BEA normally addresses these issues of definitions and methodology in 5-year "benchmark" revisions. In forecasting the true values of GDP, we make no assumptions about the changes these revisions make. The inability of the model to forecast changes that occur during benchmark revisions is a shortcoming of our work as well as that of other scholars in this field.

Commentary

Dean Croushore

It is a pleasure to discuss Richard Anderson and Charles Gascon's (2009) article on their attempt to develop a state-space model to measure potential output growth in the face of data revisions. They use the methodology of Cunningham et al. (2007) applied to real output, to see if they can develop a better measure of potential output than other researchers. Such an approach seems promising, and they develop a unique method to study the data.

This approach holds promise because many practical approaches based on standard statistical models or production functions have not proven reliable indicators of potential output. One reason these methods may fail could be that the data are revised and the methods used do not account for such revisions. By accounting for data revisions in a systematic way, the authors hope to develop an improved calculation of potential output.

However, if the potential output series is subject to breaks not easily detected for many years, this approach may not be fruitful—you simply must wait many years to determine what potential output is. The state-space method may be ideal for calculating latent variables that correspond to an observable variable subject to large data revisions, but it is not helpful for early detection of breaks in series like potential output. There is simply no getting around the fundamental fact that potential output inherently requires the use of a two-sided filter and will be tremendously imprecise at the end of the sample when only a one-sided filter can be used.

CONGRESSIONAL BUDGET OFFICE MEASURES OF POTENTIAL OUTPUT

Many economic models rely on the concept of potential output, yet it is not observable. As new data arrive over time, practitioners who need a measure of potential output for their models use various statistical procedures to revise their view of potential output. One such practitioner is the Congressional Budget Office (CBO), which has produced a measure of potential output since 1991. An examination of some of the changes in their measure of potential output over time helps illustrate some of the difficulties of using the concept.

Figure 1 shows the CBO January 1991 and January 1996 versions of potential output growth. The vertical bars indicate the dates the series were created. In the 1991 version, potential output growth rises in discrete steps over time; in the 1996 version, growth rates evolve more smoothly. In the 1996 version, there is substantial volatility in potential output growth in the 1970s and early 1980s; in the 1991 version it is smoother.

Figure 2 compares the CBO 1996 and 2001 versions. Differences in the series' volatility in the 1970s and early 1980s and growth rates in the 1990s and 2000s are substantial. For example, in 1996, the CBO thought potential output growth for 1996 was about 2 percent per year; but in 2001, they thought it was about 3 percent.

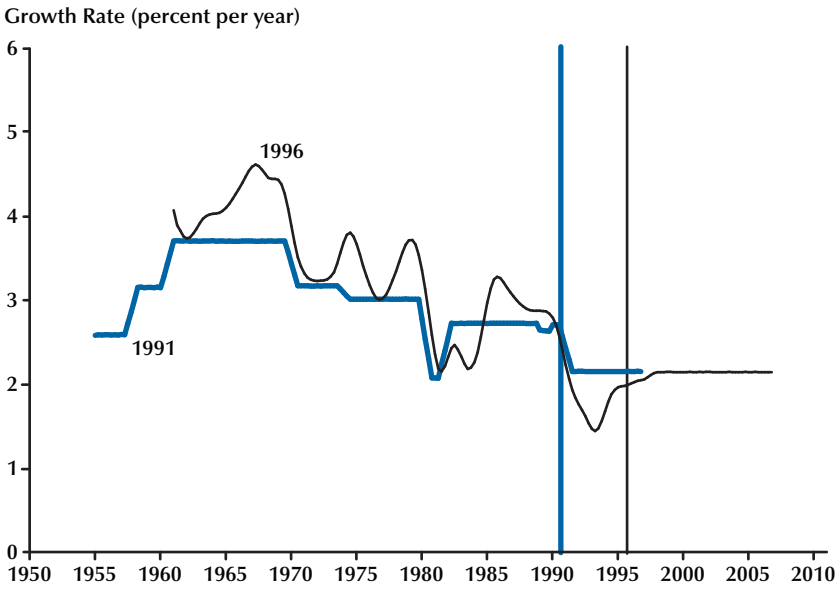
The CBO 2001 and 2008 versions of potential output growth (Figure 3) show even greater volatil-

Dean Croushore is a professor of economics and Rigsby Fellow at the University of Richmond.

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 371-81.
© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

Figure 1

CBO Potential Output Growth, 1991 and 1996

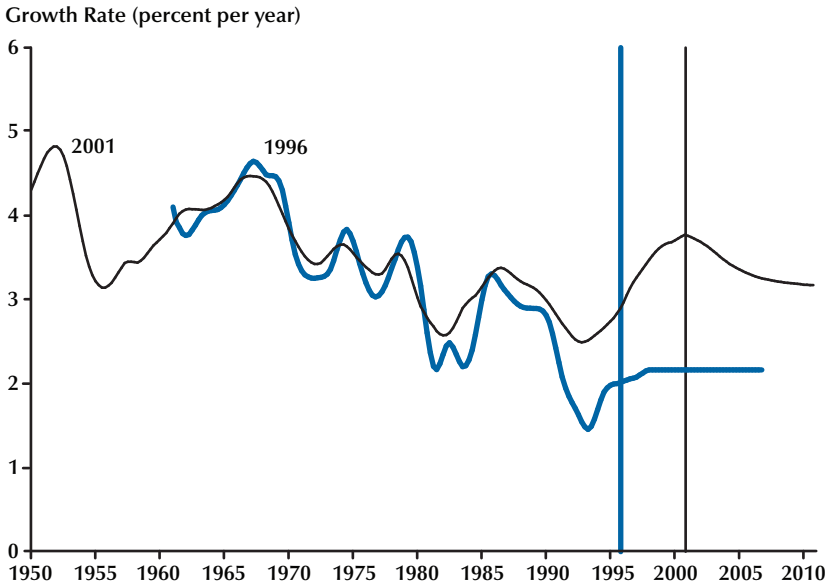


NOTE: The vertical bars indicate the dates the data series were created.

SOURCE: Federal Reserve Bank of St. Louis ArchivaL Federal Reserve Economic Data (ALFRED) database (series ID: GDPPOT).

Figure 2

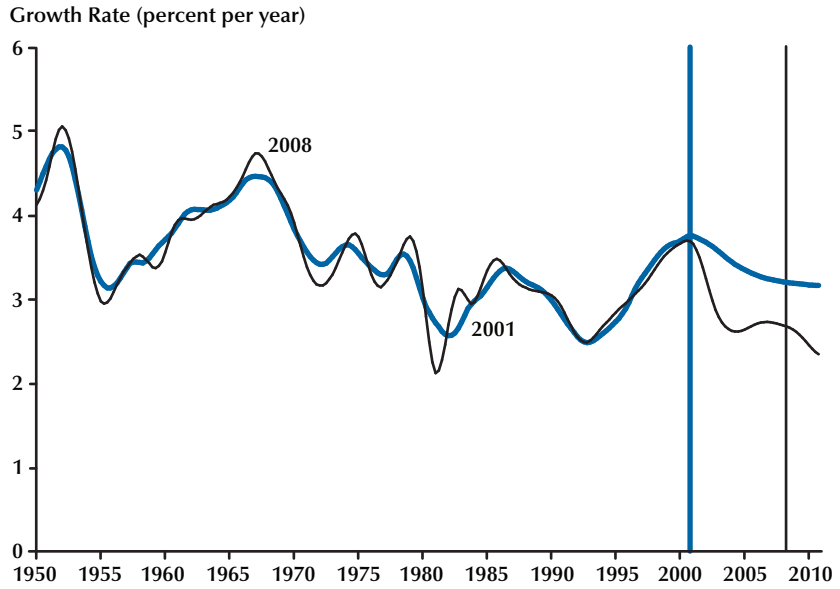
CBO Potential Output Growth, 1996 and 2001



SOURCE: Federal Reserve Bank of St. Louis ALFRED database (series ID: GDPPOT).

Figure 3

CBO Potential Output Growth, 2001 and 2008



SOURCE: Federal Reserve Bank of St. Louis ALFRED database (series ID: GDPPOT).

ity in the 1970s and early 1980s than earlier published series (see Figures 1 and 2) and a large difference in growth rates in the 2000s.

Thus, in the CBO's view, the period with the greatest revisions to potential output growth over time is the 1970s and early 1980s. In addition, in both the 1996 and 2001 versions, the potential growth rates at the end points of the sample changed substantially over time. This end-point problem is the major challenge to constructing a better measure of potential output.

KEY ASPECTS OF THE ANDERSON AND GASCON APPROACH

Key aspects of the approach taken by Anderson and Gascon include using a state-space approach (a very reasonable method) and exploiting the forecastability of data revisions following Cunningham et al. (2007). However, the real-time research literature, as described in detail in Croushore (2008a), includes few examples of macroeconomic variables whose revisions are forecastable in real time. Forecastable variables include U.S. retail sales (see Conrad and Corrado, 1979), Mexican industrial production (see Guerrero, 1993), gross domestic product (GDP) in Japan and the United Kingdom (see Faust, Roger, and Wright, 2005), and U.S. core personal consumption expenditures (PCE) inflation (see Croushore, 2008b). U.K. GDP is the focus of Cunningham et al. (2007). For U.S. GDP, revisions are not likely forecastable at all. And if this indeed is the case, the major feature of the Anderson and Gascon article could be a false trail.

A simulated out-of-sample exercise using real-time data must be performed to determine whether revisions are forecastable. Simply running a regression using an entire sample of data is not sufficient because finding a significant coefficient using the whole sample does not mean revisions are forecastable in real time.

The proper procedure to determine whether revisions are forecastable is described in Croushore

(2008b) regarding forecasting revisions to core PCE inflation. Suppose you think the initial release of data is not a good forecast of data to be released in the annual July revision of the national income and product accounts. Specifically, suppose you are standing in the second quarter of 1985 and have just received the initial release of the PCE inflation rate for 1985:Q1. You need to run a regression using as the dependent variable all the data on revisions from the initial release through the government's annual release in the current period, so the sample period is 1965:Q3–1983:Q4. So, you regress the revisions to the initial release for each date and a constant term:

$$\text{(1)} \qquad Revision(t) = \alpha + \beta\ initial(t) + \varepsilon(t).$$

Next, use the estimates of α and β to make a forecast of the August revision that will occur in 1986:

$$\hat{r}(1985\text{:Q1}) = \hat{\alpha} + \hat{\beta} \cdot initial(1985\text{:Q1}).$$

Repeat this procedure for releases from 1985:Q2 to 2006:Q4. Finally, forecast the value of the annual revision for each date from 1985:Q1 to 2006:Q4 based on the formula

$$\text{(2)} \qquad \hat{A}(t) = initial(t) + \hat{r}(t).$$

At the end of this process, examine the root mean squared forecast errors (RMSEs) as follows: Take the annual release value as the realization and compare the RMSE of the forecast of that value (given by equation (2)) with the RMSE of the forecast of that value assuming that the initial release is an optimal forecast. In such a case, the results show that it *is* possible to forecast the annual revision. Indeed, had the Federal Reserve used this procedure, it would have forecast an upward revision to core PCE inflation in 2002 and might not have worried so much about the unwelcome fall in inflation that was a major concern in this period. However, following such a method does not appear to work for U.S. real GDP. Cunningham et al. (2007) found that it worked for U.K. real GDP, but Anderson and Gascon's attempt to use it for U.S. real GDP is less likely to be fruitful. This is not to say that the initial release of U.S. real GDP data is an optimal forecast of the latest data, only that no one has successfully forecasted the revisions in the manner described above. You could argue that we should always assume that real GDP will be revised upward because the statistical agencies will always fall behind innovative processes, so GDP will be higher than initially reported. But the major reasons for upward revisions to GDP in the past include the reclassification of government spending on capital goods as investment, the change in the treatment of business software, and similar innovations that raised the entire level of real GDP. Whether similar upward revisions will occur in the future is uncertain.

THE STRUCTURE OF REAL-TIME DATA

Researchers of real-time data begin by developing a vintage matrix, consisting of the data as reported by the government statistical agency at various dates. An example is given in Table 1.

In the vintage matrix, each column represents a vintage, that is, the date on which a data series is published. For example, the first column reports the dates from 1947:Q1 to 1965:Q3 for data that would have been observable in November 1965. Each row in the matrix represents an activity date, that is, the date for which economic activity is measured. For example, the first row shows various measures for 1947:Q1. Moving across rows shows how data for a particular activity date are revised over time. The main diagonal of the matrix shows initial releases of the data for each activity date, which moves across vintages. Huge jumps in numbers indicate benchmark revisions with base-year changes. For example, in the first row, for 1947:Q1 the value rises from 306.4 in early vintages to 1570.5 in the most recent vintages.

Until about 1999, researchers studying monetary policy or forecasters building models ignored the vintage matrix and simply used the last column of the matrix available at the time—the latest data. If data revisions are small and white noise, this is a reasonable procedure. But in 1999, the Federal Reserve Bank of Philadelphia put together a large real-time dataset for macroeconomists, and it became possible for researchers and fore-

Table 1

The Vintage Matrix

	Real Output					
	Vintage (*v*)					
Activity date	11/65	02/66	05/66	…	11/07	02/08
1947:Q1	306.4	306.4	306.4	…	1,570.5	1,570.5
1947:Q2	309.0	309.0	309.0	…	1,568.7	1,568.7
1947:Q3	309.6	309.6	309.6	…	1,568.0	1,568.0
⋮	⋮	⋮	⋮		⋮	⋮
1965:Q3	609.1	613.0	613.0	…	3,214.1	3,214.1
1965:Q4	NA	621.7	624.4	…	3,291.8	3,291.8
1966:Q1	NA	NA	633.8	…	3,372.3	3,372.3
⋮	⋮	⋮	⋮		⋮	⋮
2007:Q1	NA	NA	NA	…	11,412.6	11,412.6
2007:Q2	NA	NA	NA	…	11,520.1	11,520.1
2007:Q3	NA	NA	NA	…	11,630.7	11,658.9
2007:Q4	NA	NA	NA	…	NA	11,677.4

SOURCE: Federal Reserve Bank of Philadelphia Real-Time Dataset for Macroeconomists (RTDSM; series ID: ROUTPUT).

casters to use the entire vintage matrix (see Croushore and Stark, 2001). Subsequent work at the Federal Reserve Bank of St. Louis expanded the Philadelphia Fed's work to create the vintage matrix for a much larger set of variables. The availability of such data has allowed researchers of real-time data to study data revisions and how they affect monetary policy and forecasting. The data revisions turn out to be neither small nor white noise, so accounting for data revisions is paramount.

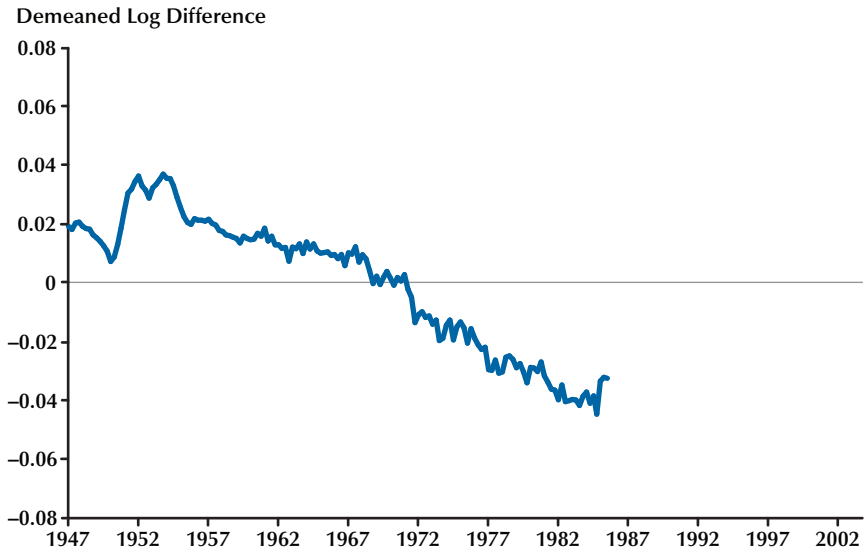
Researchers of real-time data have explored a number of ways to study what happens in the vintage matrix. One of the main distinctions in the literature that is crucial to econometric evaluation of data revisions is the distinction between "news" and "noise." Data revisions contain *news* if the initial release of the data is an optimal forecast of the later data. If so, then data revisions are not predictable. On the other hand, if data revisions reduce *noise*, then each data release equals the truth plus a measurement error; but because the data release is not an optimal forecast, it is predictable.

Empirical findings concerning news and noise are mixed. Money-supply data contain noise, according to Mankiw, Runkle, and Shapiro (1984), but GDP releases represent news, according to Mankiw and Shapiro (1986). Different releases of the same variable can vary in their news and noise content, as Mork (1987) found. For U.K. data, releases of most components of GDP contain noise, according to Patterson and Heravi (1991). The distinction between news and noise is vital to some state-space models, such as the one developed by Jacobs and van Norden (2007).

Anderson and Gascon ignore the distinction between news and noise because they develop a new and unique way to slice up the vintage matrix. Rather than focus on the vintage date, their analysis is a function of the "maturity" of data—that is, how long a piece of data for a given activity date has matured. They then track that piece of data over a length of time that they call the "revision horizon," which they can vary to discover different properties in the data of the revisions. This is a clever procedure and has the potential to lead to interesting results.

Figure 4

Stark Plot: December 1985–November 1991



SOURCE: Author's calculations using data from the Federal Reserve Bank of Philadelphia RTDSM (series ID: ROUTPUT).

The statistical model used by Anderson and Gascon is based on the following equation:

$$y_t^j = y_t + c^j + v_t^j.$$

A measured piece of data of some maturity j for activity date t is equal to the true value of the variable at activity date t, plus a bias term that is a function of maturity (but not vintage or activity date), plus a measurement error that is a function of both maturity and the activity date. This is the same method used by Cunningham et al. (2007).

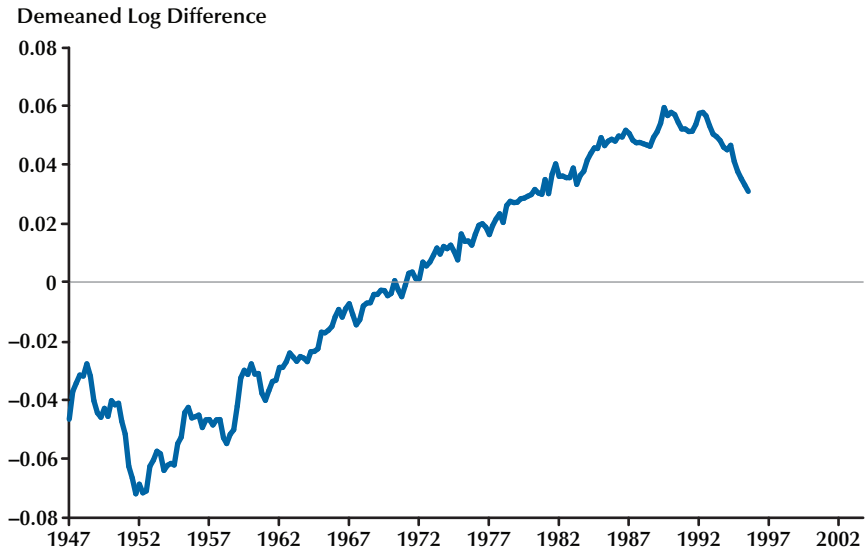
The Problem of Benchmark Revisions

Unfortunately, the Anderson and Gascon method may not work well if there are large and significant benchmark revisions to the data, because then the relationships in question would be a function of not only the activity date and maturity, but also a function of vintage, because benchmark revisions hit only one vintage of data every five years or so. But when they do hit, they affect the values of a different maturity for every activity date. So, if benchmark revisions are significant, then the Anderson and Gascon procedure could face problems.

Are benchmark revisions significant? I like to investigate the size of benchmark revisions using Stark plots, which I named after my frequent coauthor Tom Stark, who invented the plot (see Croushore and Stark, 2001). Let $X(t,s)$ represent the level of a variable that has been revised between vintages a and b, where vintage b is farther to the right in the vintage matrix and thus later in time than vintage a. Let m = the mean of $\log[X(\tau,b)/X(\tau,a)]$ for all the activity dates τ that are common to both vintages. The Stark plot is a plot of $\log[X(t,b)/X(t,a)] - m$. Such a plot would be a flat line if the new vintage were just a scaled-up version of the old one, that is, if $X(t,b) = \lambda X(t,a)$. If the plot shows an upward trend, then later data have more upward revisions than earlier data. Spikes in the plot show idiosyncratic data revisions. More important to analysis of data revisions would be any persistent deviation of the Stark plot from the zero line, which would imply a cor-

Figure 5

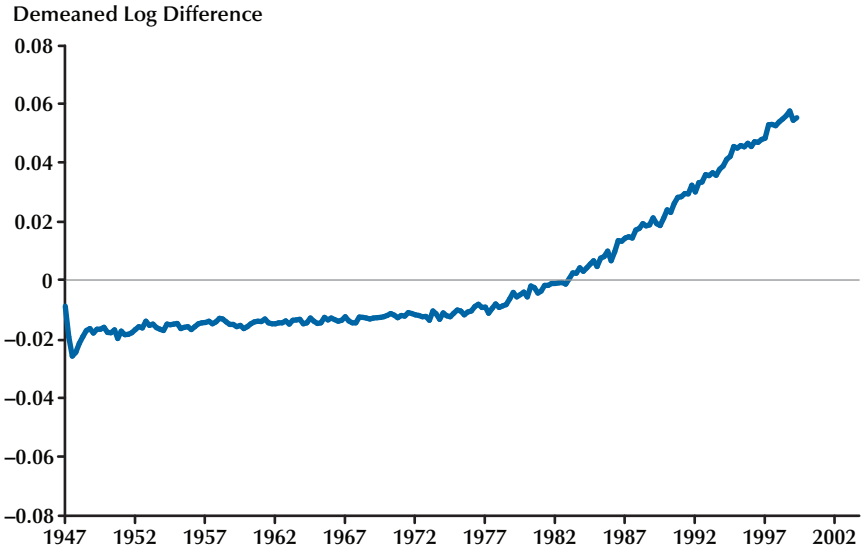
Stark Plot: December 1995–October 1999, Chain Weighting; Government Purchases Reclassified as Investment



SOURCE: Author's calculations using data from the Federal Reserve Bank of Philadelphia RTDSM (series ID: ROUTPUT).

Figure 6

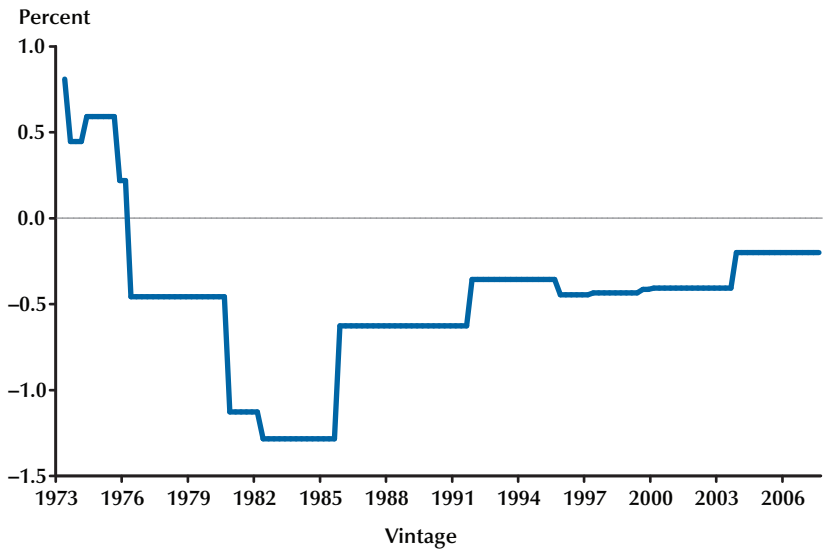
Stark Plot: October 1999–November 2003; Software Reclassified as Investment



SOURCE: Author's calculations using data from the Federal Reserve Bank of Philadelphia RTDSM (series ID: ROUTPUT).

Figure 7

Real Consumption Growth, 1973:Q2



SOURCE: Author's calculations using data from the Federal Reserve Bank of Philadelphia RTDSM (series ID: RCON).

relation of revisions arising from the benchmark revision.

In Figures 4, 5, and 6, we examine Stark plots that span particular benchmark revisions. Figure 4 shows how vintage data were revised from December 1985 to November 1991 for activity dates from 1947:Q1 to 1985:Q3. The data early in the sample period show upward revisions and those later in the sample period show downward revisions. There is a clear pattern in the data, which is mainly driven by the benchmark revision to the data that was released in late December 1985 (the December 1985 vintage date corresponds to the data as it existed in the middle of the month).

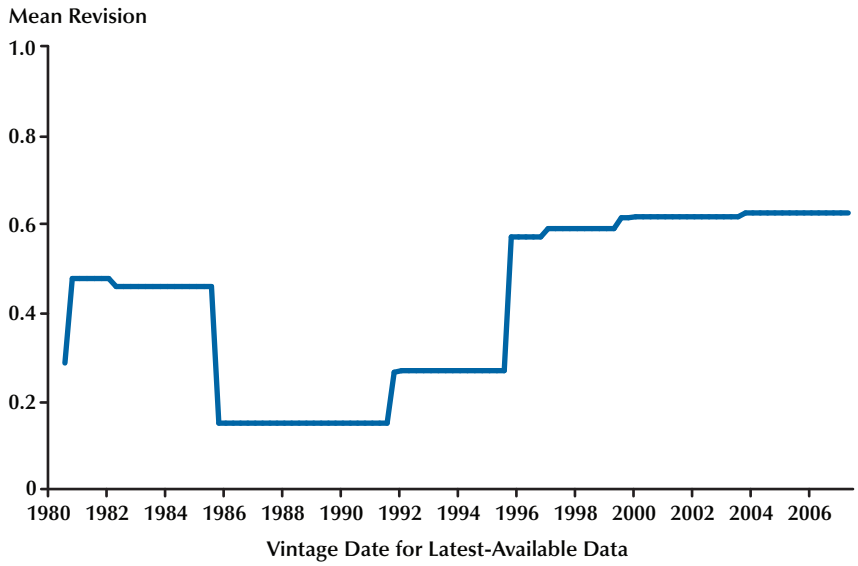
Figure 5 shows the revisions from December 1995 to October 1999, illustrating the impact of the benchmark revision of January 1996, which introduced chain weighting and reclassified government investment expenditures from their previous treatment as an investment expense subject to depreciation. The impact is very large, with data early in the sample showing downward revisions relative to data later in the sample.

Figure 6 illustrates the impact of the November 1999 benchmark revision, in which business software was reclassified as investment; we look at the changes from the October 1999 vintage to the November 2003 vintage. The nonlinear Stark plot suggests little change in growth rates in the early part of the sample, but increasing growth rates later in the sample. The impact of these changes in the benchmarks is considerable. There is clearly a significant change in the entire trajectory of the variable over time, which should be accounted for in any empirical investigation of the variable.

Do revisions ever settle down and stop occurring? In principle, they do under chain weighting, except for redefinitions that occur in benchmark revisions. For example, Figure 7 shows the growth rate of real consumption spending for activity date 1973:Q2. It has been revised by several percentage points over time and changed significantly as recently as 2003, some 30 years after the activity date. Thus, we cannot be confident that data are ever truly final and that there will never be a significant future revision.

Figure 8

Mean Revision, Initial to Latest



SOURCE: Author's calculations using data from the Federal Reserve Bank of Philadelphia RTDSM (series ID: ROUTPUT).

One key idea in the Anderson and Gascon article is to exploit the apparent bias in initial releases of the data. Unfortunately the bias seems to jump at benchmarks, as the Stark plots suggest. To see the jumps more clearly, Figure 8 plots what one would have thought the bias was at different vintage dates for real output growth. That is, it calculates the mean revision from the initial release to the latest available data, where for the sample of data from 1965:Q3 to 1975:Q3 the vintages of the latest available data are from 1980:Q3 to 2007:Q2. If we were standing in 1980:Q3, Figure 8 indicates we would have thought that the bias in the initial release of real output growth was 0.28 percentage points. But someone observing the data in the period from 1980:Q4 to 1982:Q1 would have thought it was 0.45 percentage points. And the apparent bias keeps changing over time, ending in 2007:Q2 at 0.62 percentage points. So, the bias changes depending on the date when you measure the bias. The same is true if you allow the sample period to change, rather than focusing on just one sample period as we did in Figure 8.

The Stark plots provide important information for researchers—that the bias is a function of the benchmark dates, not just maturity. Thus, equation (3) in Anderson and Gascon,

$$c^j - c^1(1+\lambda)^{j-1},$$

which treats the bias solely as maturity dependent, is not likely to work across benchmark revisions.

The other key assumption that Anderson and Gascon use in their empirical framework is that the measurement error follows an autoregressive (AR(q)) process. The Stark plots suggest that such an assumption is not well justified, because the process at different benchmark revisions is much more complicated than any AR(q) process can capture.

WHERE NEXT?

Given the issues identified here, how should the authors proceed with their research? I offer five suggestions. First, they should compare their

results on the potential output series generated by their method with that of some benchmark, such as some of the series generated in Orphanides and van Norden (2002). Second, they should examine forecasts of revisions that can be generated by the model to see if they match up reasonably well with actual revisions. Third, they should see how stable their model is when it encounters a benchmark revision. That is, if the model were used in real time to generate a series for potential output and then suddenly hit a benchmark revision, what would that do to the potential output series? Fourth, they should attempt to reconcile the Stark plots with their assumptions about the data to see how much damage such assumptions might make. Finally, because they have ignored the distinction between news and noise, they might want to consider the impact the results of Jacobs and van Norden (2007) would have on their empirical model.

CONCLUSION

The research by Anderson and Gascon is an interesting and potentially valuable contribution to estimating potential output. However, practical issues, in particular the existence of benchmark revisions, may derail it. It may be that no new empirical method can handle revisions and produce better estimates of potential output in real time than current methods. If so, then we may have to conclude that potential output cannot be measured accurately enough in real time to be of any value for policymakers.

REFERENCES

Anderson, Richard and Gascon, Charles. "Estimating U.S. Output Growth with Vintage Data in a State-Space Framework." Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 349-69.

Conrad, William and Corrado, Carol. "Application of the Kalman Filter to Revisions in Monthly Retail Sales Estimates." *Journal of Economic Dynamics and Control*, May 1979, *1*(2), pp. 177-98.

Croushore, Dean. "Frontiers of Real-Time Data Analysis." Working Paper No. 08-4, Federal Reserve Bank of Philadelphia, March 2008a; www.philadelphiafed.org/research-and-data/publications/working-papers//2008/wp08-4.pdf.

Croushore, Dean. "Revisions to PCE Inflation Measures: Implications for Monetary Policy." Working Paper 08-8, Federal Reserve Bank of Philadelphia, March 2008b.

Croushore, Dean and Stark, Thomas. "A Real-Time Data Set for Macroeconomists." *Journal of Econometrics*, November 2001, *105*(1), pp. 111-30.

Cunningham, Alastair; Eklund, Jana; Jeffery, Christopher; Kapetanios, George and Labhard, Vincent. "A State Space Approach to Extracting the Signal from Uncertain Data." Working Paper No. 336, Bank of England, November 2007.

Faust, Jon; Rogers, John H. and Wright, Jonathan H. "News and Noise in G-7 GDP Announcements." *Journal of Money, Credit, and Banking*, June 2005, *37*(3), pp. 403-19.

Guerrero, Victor M. "Combining Historical and Preliminary Information to Obtain Timely Time Series Data." *International Journal of Forecasting*, December 1993, *9*(4), pp. 477-85.

Jacobs, Jan P.A.M. and van Norden, Simon. "Modeling Data Revisions: Measurement Error and Dynamics of 'True' Values." Working paper, HEC Montreal, June 2007.

Mankiw, N. Gregory; Runkle, David E. and Shapiro, Matthew D. "Are Preliminary Announcements of the Money Stock Rational Forecasts?" *Journal of Monetary Economics*, July 1984, *14*(1), pp. 15-27.

Mankiw, N. Gregory and Shapiro, Matthew D. "News or Noise? An Analysis of GNP Revisions." *Survey of Current Business*, May 1986, *66*(5), pp. 20-25.

Mork, Knut Anton. "Ain't Behavin': Forecast Errors and Measurement Errors in Early GNP Estimates." *Journal of Business and Economic Statistics*, April 1987, *5*(2), pp. 165-75.

Orphanides, Athanasios and van Norden, Simon. "The Unreliability of Output Gap Estimates in Real Time." *Review of Economics and Statistics*, November 2002, *84*(4), pp. 569-83.

Patterson, K.D. and Heravi, S.M. "Data Revisions and the Expenditure Components of GDP." *Economic Journal*, July 1991, *101*(407), pp. 887-901.

Panel Discussion

The Role of Potential Output Growth in Monetary Policymaking in Brazil

Carlos Hamilton Araujo

Potential output is important in policymaking for a number of reasons:

- It is a key variable in most macroeconomic models because it enables construction of measures of the output gap. These measures are often used in the IS and Phillips curves and the Taylor rule, among others.
- It provides a measure of economic slack (i.e., its cyclical position).
- It helps to gauge future inflation pressures.
- It is important for estimating cyclically adjusted variables (e.g., structural fiscal deficit).

However, potential output is difficult to handle. As a latent variable, it is hard to measure in any circumstance, and frequent data revisions worsen the accuracy of any estimation. For example, in Brazil, the available time series has a short data span and the methodology for calculating gross domestic product (GDP) has changed frequently. Another potential problem is that geographic data might also be inadequate (e.g., the unemployment rate).

The Central Bank of Brazil uses a variety of statistical methods to measure potential output. The most common are statistical filters, including the Hodrick-Prescott filter, band-pass filters, Kalman filters, and Beveridge-Nelson decompositions. These methods are not based on economic theory or models, and each has its idiosyncrasies—sometimes with opposite identifying assumptions. As a general rule, the rationale is the same for all: to decompose the GDP time series into a permanent component and a transitory, cyclical component to measure the output gap. It is a shortcoming of these measures that they do not consider information other than GDP itself. They often behave like moving averages and, hence, perform poorly when the original GDP series faces large and sudden changes. In addition, the resulting filtered time series is frequently judged too volatile relative to the prior beliefs of senior policymakers.

We also use macroeconomic methods, including Cobb-Douglas production functions, structural vector autoregressions, dynamic stochastic general equilibrium models, and other macro models. To some extent, these are based on economic theory and may impose quite strong restrictions on the data. In addition, estimates are model dependent, which often leads to disagreement regarding the "true" model; furthermore, estimates are sensitive to model specification error. Given these restrictions, these models might be more difficult to estimate than with the previous statistical methods and may ignore key determinants of potential output.

Now, consider the simplest production function approach:

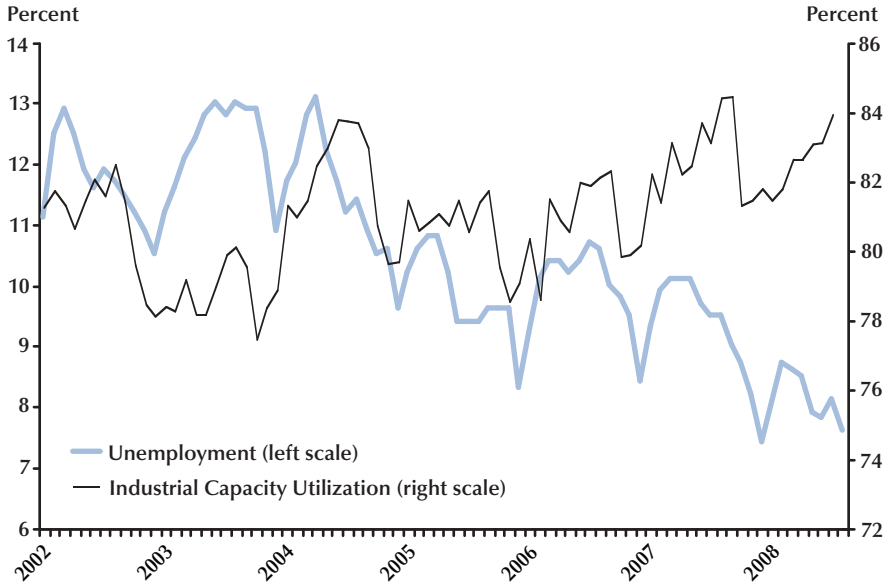
$$\bar{Y}_t = \bar{A}_t \bar{K}_t^{\alpha} \bar{L}_t^{1-\alpha}.$$

Carlos Hamilton Araujo is the head of the research department at the Central Bank of Brazil.

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 383-85.
© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, the regional Federal Reserve Banks, or the Central Bank of Brazil. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

Figure 1

Unemployment Rate and Capacity Utilization



This approach is based on widely accepted economic theory and explicitly specifies the sources of economic growth. An alternative specification that we find useful is the following:

$$Gap_t = \alpha\left[\log(cu_t) - \ln(NAIRCU_t)\right] + (1-\alpha)\left[\log(1-un_t) - \ln(1-NAIRU_t)\right],$$

where *NAIRCU* is the natural rate of capacity utilization and *NAIRU* is the natural unemployment rate. The unemployment rate and capacity utilization rate for Brazil are shown in Figure 1.

Regardless of the adopted measure, potential output estimates are always uncertain. In this sense, the Bank relies on additional economic variables as a cross-check of economic activity; these variables include unemployment, capacity utilization, industrial production, retail sales, wage growth, and surveys of corporate confidence. Thus, various potential output measures are compared by computer simulations, focused on using Phillips curves to forecast inflation and on comparisons with predictions from Okun's law. The relationship between output gap estimates and potential inflationary pressures is of utmost importance to the Monetary Policy Committee. Yet, in my view, indicators of inflation expectations are more important drivers of policy decisions than the output gap.

Potential output and capital growth are essential elements of capital deepening and output growth. Both have shown significant acceleration in recent years. Although explanations for the acceleration are uncertain, possible reasons include increased macroeconomic stability due to a new political environment (more favorable to planning); strong inflows of foreign capital in the form of foreign direct investment, bringing with it new technology; exchange rate appreciation, which sharply reduced the cost of imported capital goods; and the culmination of educational improvements, resulting in a higher-quality labor force.

Looking forward, we anticipate a slowing of economic growth. It is likely that slower GDP growth will adversely affect potential output

through similar channels: a reduction in foreign direct investment inflows; less-accommodative credit conditions, both in the domestic and international markets; and exchange rate depreciation that will increase the cost of imported capital goods.

In a nutshell: Potential output is regarded as a key indicator for assessing the slack in the economy and gauging the buildup of inflationary pressures. Because it is not observable, potential output estimates are imprecise and worsened by short and volatile time series. At the Central Bank of Brazil, we use many purely statistical and structural methods to assess potential output. Evidence and experience favor structural methods. We seek to mitigate the related uncertainties by using several methods, as well as other excess demand indicators. In policymaking, the Bank places a larger weight on inflation expectations, in addition to its estimates of the output gap.

The Role of Potential Growth in Policymaking

Seppo Honkapohja

DEFINITIONS OF POTENTIAL OUTPUT AND POTENTIAL GROWTH

Currently, differing concepts of potential output and potential growth are used in both academic research and policy discussions. Traditionally, potential output and potential growth are measures of the average productive capacity of an economy and its change over time. Correspondingly, the output gap is the deviation of actual output from its potential value, that is, from average output. If potential output is viewed as (in some sense) average output, then potential output is naturally measured by fitting a statistical trend on the path of output over time. John Taylor's (1993) seminal paper on estimated interest rate rules used such a traditional measure for the output gap. Alternatively, potential growth might be measured by fitting trends to paths of factor supplies and using these in an estimated production function.

Nowadays, it is common to use a specific economic model to estimate potential output, its growth rate, and the output gap. Here is a simple example. Consider an economy with perfect competition and a Cobb-Douglas production function, $Y_t = A_t K_t^{\alpha} N_t^{1-\alpha}$. The log-linearization of this production function is

$$y_t = \alpha k_t + (1-\alpha) n_t + a_t, \tag{1}$$

where lower-case letters denote logarithms of output, capital, labor input, and total factor productivity (TFP). The log of TFP, a_t, evolves exogenously, while the actual values of y_t and n_t are determined as part of the competitive equilibrium. The difficulty is that TFP cannot be directly observed but must be obtained as a residual from equation (1), using an estimate or calibrated value for α.[1] With such an estimate, we obtain potential output, y_t^p, as

$$y_t^p = \bar{\alpha}_t k_t + (1-\bar{\alpha}_t) n_t + \bar{a}_t,$$

where $\bar{a}_t$ and $\bar{\alpha}_t$ are estimates of TFP and parameter α, respectively, for period t. Although model-based, this calculation often produces measures

[1] As is well known, there are also more sophisticated ways to estimate TFP. For example, see Chambers (1988, Chap. 6) for an introductory discussion.

Seppo Honkapohja is a member of the Board of the Bank of Finland and a research fellow at the Centre for Economic Policy Research.

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 385-89.
© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, the regional Federal Reserve Banks, or the Bank of Finland. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

close to the traditional statistical measures. In this case, the output gap is $y_t - y_t^p$.

In the preceding formulation, $\bar{a}_t$ and $\bar{\alpha}_t$ may be ex post or real-time estimates of TFP and α, respectively. In practice, there are short- to medium-term policy concerns that require real-time measurement of the output gap. A possible policy objective can be to smooth fluctuations in aggregate output. Of course, in a competitive economy without distortions, there is no reason to offset the random fluctuations in y_t, as the equilibrium is Pareto efficient. If there are distortions, one might have some interest in measuring the real-time output gap, but this depends on the nature of the distortions and whether they vary cyclically. Naturally, measurement of potential output and potential growth is also important for setting growth policy; for example, the so-called Lisbon Agenda was devised to address the sluggish growth of most Western European Union countries. Such issues are long-run policy concerns and the background studies are based on, for example, growth accounting methodologies with ex post estimates for parameters. In such cases, there is no urgency to obtain real-time measurement for potential output.

STICKY PRICES

Though some disagreements exist, it is now a common view that the perfectly competitive, flexible-price model is not relevant for short- to medium-run policymaking. The current workhorse for monetary policy analysis is the New Keynesian (NK) model, which differs from the perfect-competition model in two crucial respects: In it the economy is imperfectly competitive and displays nominal price and/or wage rigidity.[2]

We modify the model outlined above by introducing differentiated goods and imperfect competition. Log-linearized optimal consumption behavior as a log-deviation from the steady state is described by the Euler equation,

$$c_t = E_t c_{t+1} - \sigma^{-1}\left(i_t - E_t \pi_{t+1} - \rho\right),$$

where $\rho = -\log\beta$ and β is the subjective discount factor of the economy and σ is a utility function parameter. In equilibrium,

$$C_t = Y_t$$

and, therefore, we obtain the dynamic IS curve,

$$y_t = E_t y_{t+1} - \sigma^{-1}\left(i_t - E_t \pi_{t+1} - \rho\right). \tag{2}$$

Here lower-case variables denote log-deviations from the steady state. Equation (2) indicates that aggregate output in the economy depends positively on expectations of next-period output and negatively on the real rate of interest, where the latter is defined in terms of expected next-period inflation.

The dynamics of inflation are described by an aggregate supply curve, also called the NK Phillips curve:

$$\pi_t = \beta E_t \pi_{t+1} + \lambda\left(mc_t - mc\right),$$

where inflation (as a deviation from the steady state) depends on expected inflation and the deviation of marginal cost from its steady-state value. Here λ is a function of several structural parameters. It can be shown that

$$mc_t = w_t - p_t - \frac{1}{1-\alpha}\left(a_t - \alpha y_t\right) - \log\left(1-\alpha\right).$$

It is possible to write $mc_t - mc$ in terms of a new measure of the output gap:

$$\tilde{y}_t = y_t - y_t^n, \text{ where}$$

$$y_t^n = \frac{1+\psi}{\sigma(1-\alpha)+\psi+\alpha} a_t - \frac{(1-\alpha)\left(\mu - \log(1-\alpha)\right)}{\sigma(1-\alpha)+\psi+\alpha}.$$

Here y_t^n is the natural level of output, that is, aggregate output at the flexible price (but monopolistically competitive) level.[3] Note that $y_t^n < y_t^{CE}$ because of imperfect competition. Note also that the natural level of output is different from potential output of the economy.

[2] There are several good expositions of the NK model. The formal details below are based on the excellent exposition of the NK model in Galí (2008).

[3] ω is a utility function parameter, whereas κ is the log of the steady-state markup.

Using the output gap, the inflation equation can be written

$$\pi_t = \beta E_t \pi_{t+1} + \kappa \tilde{y}_t,$$

which implies that from a business cycle viewpoint the output gap, measured as just explained, is the relevant concept for monetary policy analysis. The dynamic IS curve can also be written in terms of the output gap as

$$\tilde{y}_t = E_t \tilde{y}_{t+1} - \sigma^{-1}\left(i_t - E_t \pi_{t+1} - r_t^n\right), \text{ where}$$

$$r_t^n = \rho + \sigma \frac{1+\psi}{\sigma(1-\alpha)+\psi+\alpha} E_t\left(\Delta a_{t+1}\right).$$

Here r_t^n is the natural rate of interest.

To summarize, monetary policy analysis uses two different concepts of the output gap and both are used in monetary policy analysis. The traditional concept of potential output and the output gap are defined by the deviation from trend, whereas the recent model-based notion of the output gap is defined as the difference between actual output and the flexible price level of output. The two concepts are different and can behave in different ways, as vividly illustrated by Edge, Kiley, and Laforte (2008, Figure 1) and also studied by, for example, Andres, Lopez-Salido, and Nelson (2005) and Justiniano and Primiceri (2008).

NOISY DATA

The NK model outlined above suggests that, in theory, the output gap measure $\tilde{y}_t$, derived in the NK model (or analogously in dynamic stochastic general equilibrium models), is the appropriate measure of potential output for monetary policy analysis. It should be emphasized that this view holds only in theory for several reasons. First, any model-based output-gap measure is model dependent and thus capable of generating misleading recommendations. One should always examine the robustness of conclusions based on a specific model and the corresponding measure.

Second, how a measure will be used should be considered before deciding which model to use. The output gap measure based on the NK model is intended for analysis of inflation control and often does not measure well the economy's deviations from its long-term productive capacity.

Third, even if one opts for the measure based on the NK model, assumptions about the availability of output gap information are very strong in the standard analysis of monetary policy in dynamic stochastic general equilibrium models. Policies that perform well under the usual rational expectations (RE) assumption, for example, often do not perform well if the measurement of the output gap and other variables contain significant noise. Orphanides emphasizes this problem in a number of papers (e.g., see Orphanides, 2003). In particular, he states that naive optimal policies derived under RE often do poorly if there are noisy measurements of the true variables.[4]

In principle, optimal control policies that take into account the measurement problem can be calculated using Kalman filters; however, this approach can be sensitive to measurement problems caused by imperfect knowledge. Neither the "correct model" nor the data are, in practice, fully known to the policymaker (further discussed below). The use of well-performing simple rules offers another approach to the problem of noisy measurements. In some cases, though not optimal, simple rules work better than naive optimal rules. Such simple rules have the same functional form as naive optimal rules but respond to noisy real-time data appropriately when the policy coefficients are chosen optimally.

OTHER ASPECTS OF IMPERFECT KNOWLEDGE

Noisy data are just one aspect of knowledge imperfections that policymakers face, and although there are several others, I focus on learning effects—that is, that economic evolution in the short to medium run can be significantly influenced by learning effects from economic agents trying to improve their knowledge. The literature on learning and macroeconomics has been widely researched in recent years, and mone-

[4] In addition, data revisions are often significant and make it difficult to use model-based measures, so they will not be discussed further.

tary policy design has been shown to be affected by one's learning viewpoint.

The basic ideas in learning are that (i) agents and policymakers have imperfect knowledge, (ii) expectations are based on existing knowledge and updated over time using econometric techniques, and (iii) expectations feed into decisions by agents and hence to actual outcomes and future forecasts. Learning dynamics converge to an RE equilibrium, provided that the economy satisfies an expectational stability criterion. Good policy facilitates convergence of learning.

Basic learning models use fairly strong assumptions: (i) Functional forms of agents' forecasting models are correctly specified relative to the RE equilibrium, (ii) agents accurately observe relevant variables, and (iii) economic agents trust their forecasting model. Most of these assumptions have been weakened in the recent literature. Misspecification is certainly one concern because it can inhibit convergence to an RE equilibrium and create a restricted-perceptions equilibrium. However, the implications of this for policy design are not further discussed here.

Noisy measurements have been incorporated into some models of monetary policy that include learning, most notably by Orphanides and Williams (2007 and forthcoming). Basically, these models show that the ideas discussed above still hold. One can try to consider filtering and learning together, but this is likely to be formally demanding and has not been studied. Alternatively, one can use simple rules that work well. In particular, the recent papers by Orphanides and Williams (2007, forthcoming) suggest the use of rules that do not rely on data subject to significant noise.

A specific measurement problem is agents' private expectations. It has been shown that expectations-based optimal rules would work well for optimal monetary policy design. If there are significant errors in measuring private-sector expectations, one can try to develop proxies for them. This is, in fact, typically done, perhaps using survey data from either professional forecasters or consumer surveys. An alternative is model-based proxies from a variety of sources, including indexed and non-indexed bonds, swaps, or information from purely statistical forecasting models.

If agents do not trust their personal forecasting model, then they may wish to allow for uncertainty in their forecasting model and/or their behavioral attitudes. If one allows for unspecified model uncertainty in estimation, then robust estimation methods can be used. In fact, a "maximally robust" estimation leads to so-called constant-gain stochastic gradient algorithms, which have been studied for learning in Evans, Honkapohja, and Williams (forthcoming). Of course, literature on economic behavior in the presence of unstructured model uncertainty abounds (see Hansen and Sargent, 2007). In policy design, one can also incorporate aspects of robust policy with respect to learning by private agents. Usually, it is assumed that the policymaker does not know the learning rules of private agents,[5] but considers as policy constraints *E*-stability conditions for private-agent learning; that is, recursive least squares learning is assumed. One could make additional assumptions about learning or even identify stability conditions that are robust in some sense (see, e.g., Tetlow and von zur Muehlen, 2009).

REFERENCES

Andres, Javier; Lopez-Salido, J. David and Nelson, Edward. "Sticky-Price Models and the Natural Rate Hypothesis." *Journal of Monetary Economics*, July 2005, *52*(5), pp. 1025-53.

Chambers, Robert G. *Applied Production Analysis*. Cambridge: Cambridge University Press, 1988.

Edge, Rochelle M.; Kiley, Michael T. and Laforte, Jean-Philippe. "Natural Rate Measures in an Estimated DSGE Model of the U.S. Economy." *Journal of Economic Dynamics and Control*, August 2008, *32*(8), pp. 2512-35.

Evans, George W.; Honkapohja, Seppo and Williams, Noah. "Generalized Stochastic Gradient Learning." *International Economic Review* (forthcoming).

[5] A few papers consider policy optimization with respect to learning rules of private agents.

Galí, Jordi. *Monetary Policy, Inflation, and the Business Cycle*. Princeton: Princeton University Press, 2008.

Hansen, Lars Peter and Sargent, Thomas J. *Robustness*. Princeton: Princeton University Press, 2007.

Justiniano, Alejandro and Primiceri, Giorgio E. "Potential and Natural Output." Unpublished manuscript, Northwestern University, June 2008; http://faculty.wcas.northwestern.edu/~gep575/JPgap8_gt.pdf.

Orphanides, Athanasios. "Monetary Policy Evaluation with Noisy Information." *Journal of Monetary Economics*, April 2003, *50*(3), pp. 605-31.

Orphanides, Athanasios and Williams, John C. "Robust Monetary Policy with Imperfect Knowledge." *Journal of Monetary Economics*, July 2007, *54*(5), pp. 1406-35.

Orphanides, Athanasios and Williams, John C. "Monetary Policy Mistakes and the Evolution of Inflation Expectations," in Michael D. Bordo and Athanasios Orphanides, eds., *The Great Inflation*. Chicago: University of Chicago Press (forthcoming).

Taylor, John B. "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy*, December 1993, *39*, pp. 195-214.

Tetlow, Robert and von zur Muehlen, Peter. "Robustifying Learnability." *Journal of Economic Dynamics and Control*, February 2009, *33*(2), pp. 292-316.

The Role of Potential Output in Policymaking*

James Bullard

Often, economists equate potential output with the trend in real gross domestic product (GDP) growth. My discussion is focused on "proper" detrending of aggregate data. I will emphasize the idea that theory is needed to satisfactorily detrend data—explicit theory that encompasses simultaneously both longer-run growth and shorter-run fluctuations. The point of view I wish to explore stresses that both growth *and* fluctuations must be included in the same theoretical construct if data are to be properly detrended. Common atheoretic statistical methods are not acceptable. When detrending data, an economist should detrend by the theoretical growth path so as to correctly distinguish output variance in the model due to growth from the variation in the model due to cyclical fluctuations.

The quest to fully integrate growth and cycle was Prescott's initial ambition; however, it is difficult to develop a model that can match the curvy, time-varying growth path often envisioned as describing an economy's long-run development. Instead, the Hodrick-Prescott (HP) filter was proposed to remove from the data a flexible time-varying trend (Hodrick and Prescott, 1980). My argument is that this procedure is unsatisfactory.

The idea in question is: How can we specify a model that will make the growth path look like the ones we see in the data? My suggestion is that,

* This discussion is based on the panel discussion, "The Role of Potential Output in Policymaking," available at http://research.stlouisfed.org/econ/bullard/FallPolicyConferenceBullard16oct2008.pdf and on Bullard and Duffy (2004).

James Bullard is president of the Federal Reserve Bank of St. Louis.

Federal Reserve Bank of St. Louis *Review*, July/August 2009, *91*(4), pp. 389-95.
© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the FOMC. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

as an initial approach, we use a mainstream core growth model augmented with occasional trend breaks and learning. Learning helps the model fit the data and has important implications for policy analysis. I will discuss some applications of this idea in Real Business Cycle (RBC) and New Keynesian (NK) models from Bullard and Duffy (2004) and Bullard and Eusepi (2005).

MAIN IDEAS

The equilibrium business cycle literature encompasses a wide class of models, including RBC, NK, and multisector growth models. Various frictions can be introduced in all of these approaches. Many analyses do not include any specific reference to growth, but all are based on the concept of a balanced growth path.

I will focus on a framework that is very close to the RBC model. This will provide a well-understood benchmark. However, I stress that these ideas have wide applicability in other models as well, and I will briefly discuss an NK application at the end.

Empirical studies, such as Perron (1989) and Hansen (2001), have suggested breaks in the trend growth of U.S. economic activity. One reasonable characterization of the data is to assume log-linear trends with occasional trend breaks but no discontinuous jumps in level—that is, a linear spline. This is the bottom line of Perron's econometric work. These breaks, however, suggest a degree of nonstationarity that is difficult to reconcile with available theoretical models. This is where adding learning can be helpful.

CURRENT PRACTICE

The standard approach in macroeconomics today is to analyze separately business cycles and long-term growth. The core of this analysis is the statistical trend-cycle decomposition. The standard method in the literature for trend-cycle decomposition is to use atheoretic, univariate statistical filters, that is, to conduct the decomposition series by series (see, for example, King and Rebelo, 1999). This method ignores an important dictum implied by the balanced growth path assumption: There are restrictions as to how the model's variables can grow through time and in turn, therefore, how one is allowed to detrend data. In an appalling lack of discipline, economists ignore this dictum and detrend data individually, series by series, which makes little sense in any growth theory. An acceptable theory specifies growth paths for the model's variables (i.e., consumption, investment, output); individual trends should not be taken out of the data. Still, the ad hoc practice dominates the literature.

Most of my criticisms are well known:

- Statistical filters do not remove the "trend" that the balanced growth path requires.
- Current practice does not respect the cointegration of the variables, that is, the multivariate trend that the model implies.
- Filtered trends imply changes in growth rates over time; agents would want to react to these changes and adjust their behavior. A model without growth does not allow for this change in behavior.
- The "business cycle facts" are not independent of the statistical filter employed. The econometrics literature normally—but not always—filters the data so as to achieve stationarity for estimation and inference without regard to the underlying theory's balanced growth assumption. Even recent sophisticated models (for examples, Smets and Wouters, 2007) do not address this issue.

HOW TO IMPROVE ON THIS?

The criticisms are correct in principle. They are quantitatively important. And, these issues cannot be resolved by using alternative statistical filters, because those filters are atheoretic. Instead, theory should be used to tell us what the growth path should look like; then, this theoretical trend can be used to detrend the data.

In the model I discuss, agents are allowed to react to trend changes. The ability to react to changes in trends alters agents' behavior—how much they save, how much they consume, and

so forth. Of course, this is demanding territory. I am insisting that the theorist specify both the longer-term growth and short-run business cycle aspects of a model, and then explain the model's coherence to observed data. This is the research agenda I propose.

Core Ideas

The core idea is that modelers should use "model-consistent detrending," that is, the trends that are removed from the data are the same as the trends implied by the specified model. Presumably, changes in trend are infrequent and, perhaps with some lag, are recognized by agents who then react to them. This suggests a role for learning. In addition, the cointegration of the variables or the different trends in the various variables implied by the balanced growth path is respected.

FEATURES OF THE ENVIRONMENT

As an example, I will discuss briefly the most basic equilibrium business cycle model with exogenous stochastic growth, but replace rational expectations with learning as in Evans and Honkapohja (2001). This model perhaps is appropriate when there is an unanticipated, rare break in the trend (for example, a labor productivity slowdown or acceleration). I assume agents possess a tracking algorithm and are able to anticipate the characteristics of the new balanced growth path that will prevail after the productivity slowdown occurs. If there is no trend break for a sufficient period, then there is convergence to the rational expectations equilibrium associated with that balanced growth path. Learning helps around points where there is a structural break of some type by allowing the economy to converge to the new balanced growth path following the structural break. In order for this to work, of course, the model must be expectationally stable such that the model's implied stochastic process will remain near the growth path.

Environment

The environment studied by Bullard and Duffy (2004) is a standard equilibrium business cycle model such as the one studied by Cooley and Prescott (1995) or King and Rebelo (1999). A representative household maximizes utility defined over consumption and leisure. Physical capital is the only asset. Business cycles are driven by shocks to technology. Bullard and Duffy (2004) include explicit growth in the model. Growth in aggregate output is driven by exogenous improvements in technology over time and labor force growth. The growth rate is exogenous and constant, except for the rare trend breaks that are incorporated. The production technology is standard. Under these assumptions, aggregate output, consumption, investment, and capital will all grow at the same rate along a balanced growth path.

Structural Change

The idea of structural change in this setting is simply that either the growth rate of technology or of the labor force takes on a new value. In the model, changes of this type are unanticipated. This will dictate a new balanced growth path, and the agents learn this new balanced growth path.

In order to use the learning apparatus as in Evans and Honkapohja (2001), a linear approximation is needed. Using logarithmic deviations from steady state, one can define and rewrite the system appropriately, as Bullard and Duffy (2004) discuss extensively. One must be careful about this transformation because the steady-state values can be inferred from some types of linear approximations, but we really don't want to inform the agents that the steady state of the system has changed. We want the agents to be uncertain where the balanced growth path is and learn the path over time.

Recursive Learning

Bullard and Duffy (2004) study this system under a recursive learning assumption as in Evans and Honkapohja (2001). They assume agents have no specific knowledge of the economy in which they operate, but are endowed with a perceived law of motion (PLM) and are able to use this PLM—a vector autoregression—to learn the rational expectations equilibrium. The rational expectations equilibrium of the system is deter-

minate under the given parameterizations of the model.

Should a trend break occur—say, a productivity slowdown or speedup—the change will be manifest in the coefficients associated with the rational expectations equilibrium of this system. The coefficients will change; agents will then update the coefficients in their corresponding regressions, eventually learning the correct coefficients. These will be the coefficients that correspond to the rational expectations equilibrium after the structural change has occurred.

Expectational Stability

For this to work properly the system must be expectationally stable. Agents form expectations that affect actual outcomes; these actual outcomes feed back into expectations. This process must converge so that, once a structural change occurs, we can expect the agents to locate the new balanced growth path. Expectational stability (*E*-stability) is determined by the stability of a corresponding matrix differential equation, as discussed extensively by Evans and Honkapohja (2001). A particular minimal state variable (MSV) solution is *E*-stable if the MSV fixed point of the differential equation is locally asymptotically stable at that point. Bullard and Duffy (2004) calculated *E*-stability conditions for this model and found that *E*-stability holds at baseline parameter values (including the various values of technology and labor force growth used).

WHAT THE MODEL DOES

The description above yields an entire system—one possible growth theory along with a business cycle theory laid on top of that. A simulation of the model will yield growing output and growing consumption, and so on, but at an uneven trend rate depending on when the trend shocks occur and how fast the learning guides the economy to the new balanced growth path following such a shock. The data produced by the model look closer to the raw data we obtain on the economy, and now we would like to somehow match up simulated data with actual data.

Of course, this model is too simple to match directly with the data, but it is also a well-known benchmark model so it is possible to assess how important structural change is when determining the nature of the business cycle as well as for the performance of the model relative to the data.

One aspect of this approach is that the model provides a global theory of the whole picture of the data. The components of the data have to add up to total output. This is because in the model it adds up and one is using that fact to detrend across all of the different variables. When considering the U.S. data, then, one has to think about the pieces that are not part of the model and how those might match up to objects inside the model. Bullard and Duffy (2004) discuss this extensively.

Breaks Along the Balanced Growth Path

The slowdown in measured productivity growth in the U.S. economy beginning sometime in the late 1960s or early 1970s is well known, and econometric evidence on this question is reviewed in Hansen (2001). Perron (1989) associated the 1973 slowdown with the oil price shock. The analysis by Bai, Lumsdaine, and Stock (1998) suggests the trend break most likely occurred in 1969:Q1.

The Bullard and Duffy (2004) model says that the nature of the balanced growth path—the trend—is dictated by increases in productivity units $X(t)$ and increases in the labor input $N(t)$. To find break dates, instead of relying on econometric evidence alone, Bullard and Duffy (2004) designed an algorithm that uses a simulated method of moments search process (genetic algorithm)[1] to choose break dates for the growth factors and the growth rates of these factors, based on the principle that the trend in measured productivity and hours from the model should match the trend in measured productivity and hours from the data. Table 1 reports their findings. The algorithm suggests one trend break date in the early 1960s for the labor input and two break dates for productivity: one in the early 1970s and one in the 1990s.

[1] See Appendix B in Bullard and Duffy (2004).

Table 1

Optimal Trend Breaks

	$N(t)$	$X(t)$
Initial annual growth rate (percent)	1.20	2.47
Break date	1961:Q2	1973:Q3
Mid-sample annual growth rate (percent)	1.91	1.21
Break date	—	1993:Q3
Ending annual growth rate (percent)	1.91	1.86

Table 2

Business Cycle Statistics, Model-Consistent Detrending

	Volatility		Relative volatility		Contemporaneous correlations	
	Data	Model	Data	Model	Data	Model
Output	3.25	3.50	1.00	1.00	1.00	1.00
Consumption	3.40	2.16	1.05	0.62	0.60	0.75
Investment	14.80	8.86	4.57	2.53	0.65	0.92
Hours	2.62	1.54	0.81	0.44	0.65	0.80
Productivity	2.52	2.44	0.77	0.70	0.61	0.92

According to Table 1, productivity grows rapidly early in the sample, then slowly from the ’70s to the ’90s and then somewhat faster after 1993. After each one of those breaks the agents in the model are somewhat surprised, but their tracking algorithm allows them to find the new balanced growth path that is implied by the new growth rates.

This model includes both a trend and a cycle. Looking at the simulated data from the model, what would a trend be? A trend is the economy’s path if only low-frequency shocks occur. Bullard and Duffy (2004) turn off the noise on the business cycle shock and just trace out the evolution of the economy if only the low-frequency breaks in technology and labor force growth occur. Importantly, the multivariate trend defined this way is then the same one that is removed from the actual data. In this sense, the model and the data are treated symmetrically: The growth theory that is used to design the model is dictating the trends that are removed from the actual data.

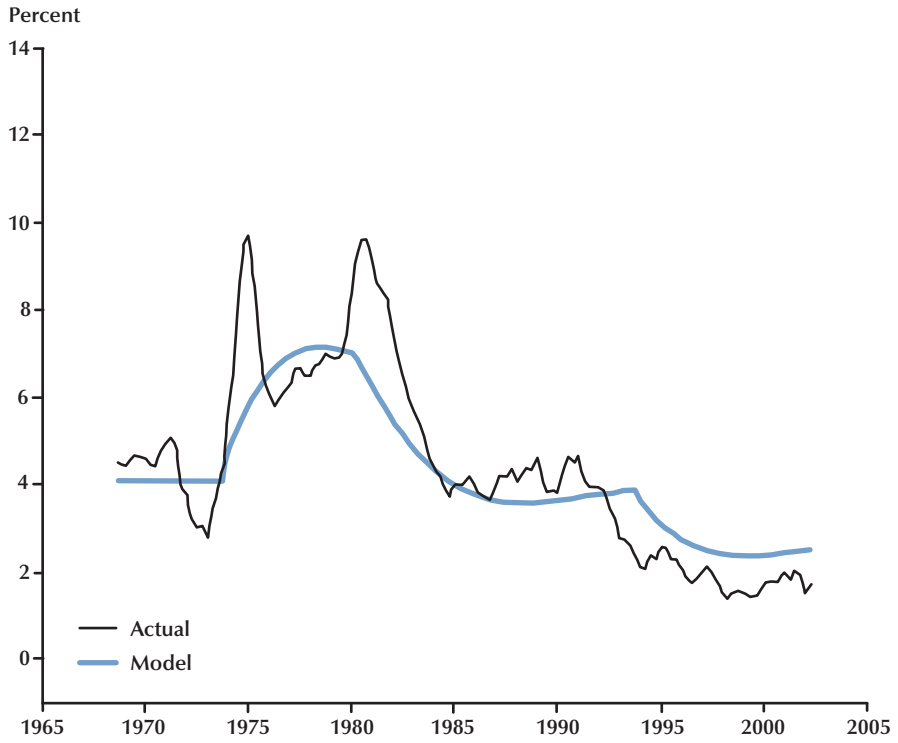
Business Cycle Statistics

The reaction of the economy to changes in the balanced growth path will depend in part on what business cycle shocks occur in tandem with the growth rate changes. Bullard and Duffy (2004) average over a large number of economies to calculate business cycle statistics for artificial economies. They collect 217 quarters of data for each economy, with trends breaking as described above. They detrend the actual data using the same (multivariate) trend that is used for the model data.

The numbers in Table 2 are not the standard ones for this type of exercise. In fact, they are quite different from the ones that are typically reported for this model, both for the data and for the model relative to the data. This shows that the issues of the underlying growth theory and its implications for the trends we expect to observe are key issues in assessing theories. One simple message from Table 2 is we obtain almost twice as much volatil-

Figure 1

U.S. Core PCE Inflation



ity in this model as there would be in the standard business cycle in this economy. This is so even though the technology shock is calibrated in the standard way.

New Keynesian Application

A similar approach can be used in the NK model. This was done by Bullard and Eusepi (2005). In the NK model (with capital), a monetary authority plays an important role in the economy's equilibrium. In Bullard and Eusepi (2005), the monetary authority follows a Taylor-type policy rule. The trend breaks and the underlying growth theory are the same as in Bullard and Duffy (2004). Now, however, one can ask how the policymaker responds using the Taylor rule given a productivity slowdown that must be learned. The policymaker initially misperceives how big the output gap is and this is making policy set the interest rate too low, pushing the inflation rate up. How large is this effect? According to Bullard and Eusepi (2005), the effect is about 300 basis points on the inflation rate for a productivity slowdown of the magnitude experienced in the 1970s (Figure 1). So, this does not explain all of the inflation in the 1970s but it helps explain a big part of it.

CONCLUSION

The approach outlined above provides some microfoundations for the largely atheoretical practices that are currently used in the literature. Structural change is not a small matter, and structural breaks likely account for a large fraction of the observed variability of output. One way to think of structural change is as a series of piecewise balanced growth paths. Learning is a glue that can hold together these piecewise paths.

I think this is an interesting approach and I would like to encourage more research that goes in this direction. It doesn't have to be a simple RBC-type model; one could instead use a more elaborate model that incorporates more empirically realistic ideas about what is driving growth and what is driving the business cycle. The approach I have outlined forces the researcher to lay out a growth theory, which is a tough and rather intensive task, but also leads to a more satisfactory detrending method and a model that is congruent with the macroeconomic data in a broad way.

REFERENCES

Bai, Jushan; Lumsdaine, Robin L. and Stock, James H. "Testing for and Dating Common Breaks in Multivariate Time Series." *Review of Economic Studies*, July 1998, *65*(3), pp. 395-432.

Bullard, James and Duffy, John. "Learning and Structural Change in Macroeconomic Data." Working Paper No. 2004-016A, Federal Reserve Bank of St. Louis, August 2004; http://research.stlouisfed.org/wp/2004/2004-016.pdf.

Bullard, James and Eusepi, Stefano. "Did the Great Inflation Occur Despite Policymaker Commitment to a Taylor Rule?" *Review of Economic Dynamics*, April 2005, *8*(2), pp. 324-59.

Cooley, Thomas F. and Prescott, Edward C. "Economic Growth and Business Cycles," in T.F. Cooley, ed., *Frontiers of Business Cycle Research*. Princeton, NJ: Princeton University Press, 1995, pp. 1-38.

Evans, George W. and Honkapohja, Seppo. *Learning and Expectations in Macroeconomics*. Princeton, NJ: Princeton University Press, 2001.

Hansen, Bruce E. "The New Econometrics of Structural Change: Dating Breaks in U.S. Labour Productivity." *Journal of Economic Perspectives*, Fall 2001, *15*(4), pp. 117-28.

Hodrick, Robert J. and Prescott, Edward C. "Postwar U.S. Business Cycles: An Empirical Investigation." Discussion Paper 451, Carnegie-Mellon University, May 1980.

King, Robert G. and Rebelo, Sergio T. "Resuscitating Real Business Cycles," in J.B. Taylor and M. Woodford, eds., *Handbook of Macroeconomics*. Volume 1C. Chap. 14. Amseterdam: Elsevier, 1999, pp. 927-1007.

Orphanides, Athanasios and Williams, John C. "The Decline of Activist Stabilization Policy: Natural Rate Misperceptions, Learning and Expectations." *Journal of Economic Dynamics and Control*, November 2005, *29*(11), pp. 1927-50.

Perron, Pierre. "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, November 1989, *57*(6), pp. 1361-401.

Smets, Frank and Wouters, Rafael. "Shocks and Frictions in U.S. Business Cycles: A Bayesian DSGE Approach." CEPR Discussion Paper No. 6112, Centre for Economic Policy Research, February 2007; www.cepr.org/pubs/dps/DP6112.asp.

Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166-0442