OLIVIER COIBION
University of Texas at Austin

YURIY GORODNICHENKO
University of California, Berkeley

MAURICIO ULATE
University of California, Berkeley

The Cyclical Sensitivity in Estimates of Potential Output

ABSTRACT The fact that declines in output since the Great Recession have been parlayed into equivalent declines in measures of potential output is commonly interpreted as implying that output will not return to previous trends. We show that real-time estimates of potential output for the United States and other countries respond gradually and similarly to both transitory and permanent shocks to output. Observing revisions in measures of potential output therefore tells us little about whether changes in actual output will be permanent. Some alternative methodologies to estimate potential output can avoid these shortcomings. These approaches suggest a much more limited decline in potential output since the Great Recession.

The Great Recession was characterized not only by large declines in economic activity in most advanced economies but also by ones that have persisted for a decade, with no sign of these affected economies catching up to previously expected trend levels. If anything, trends are now being revised down in light of these economies' continuing inability to close the output gaps first generated in 2008. As illustrated in figure 1 for the United States (see below, in section I), estimates of potential output

Conflict of Interest Disclosure: The authors did not receive financial support from any firm or person for this paper or from any firm or person with a financial or political interest in this paper. They are currently not officers, directors, or board members of any organization with an interest in this paper. No outside party had the right to review this paper before circulation.

have been systematically revised downward since the Great Recession, such that all the current deviations of output from past estimates of potential are now being reinterpreted as permanent declines in the economy's productive capacity. These large downward revisions imply that the output gap appears closed, and this absence of any remaining slackness in the economy is a primary motivation for the Federal Reserve's progressive tightening of monetary policy.

However, before we take these dynamics in the estimates of potential output at face value, we should understand their properties and what determines revisions in these estimates. In this paper, we focus on how real-time estimates of potential output respond to different economic shocks in the United States, and also across a wide range of countries. Using a variety of institutional sources for estimates of potential GDP, we find that real-time estimates of this variable respond to cyclical shocks that have no long-run effects on the economy and underrespond to shocks that do. In all cases, adjustments in real-time estimates of potential GDP are extremely gradual, much like a moving average of past output changes. In fact, given their gradual pace of adjustment to shocks and the fact that these real-time estimates fail to differentiate between shocks that do and do not affect the productive capacity of the economy, there seems to be little value added in estimates of potential GDP relative to simple measures of statistical trends. At a minimum, the fact that estimates of potential GDP are revised, either upward or downward, should *not* be taken as a sign that future changes in GDP will in fact be more or less persistent than usual but rather indicates little more than that the prior changes in GDP have been persistent.

Because estimates of potential GDP are not necessarily created in the same fashion across institutions, we consider estimates from the Federal Reserve Board and from the Congressional Budget Office (CBO) for the United States as well as estimates from the International Monetary Fund and the Organization for Economic Cooperation and Development (OECD) for a broader cross section of countries. We complement this with long-term forecasts of output growth from the professional forecasting firm Consensus Economics. Most public or international organizations follow production function approaches, in which estimates of the potential productive capacity of an economy reflect estimates of the capital stock, potential labor force sizes combined with estimates of human capital, and measures of total factor productivity (TFP). Hence, estimates of potential output should change when the technological capacity of the economy improves but not in response to purely cyclical variations in employment, such as those arising from monetary policies.

To test these propositions, we bring to bear not just a wide range of estimates of potential output but also a range of shock measures. Somewhat surprisingly, given the short samples, we find several clear patterns in the data that should give one pause before interpreting changes in estimates of potential output as indicators of permanent changes in output. First, and perhaps most strikingly, though we reproduce the common and well-documented finding that monetary shocks have only transitory effects on GDP, we then document the startling feature that these shocks are followed by a gradual change in estimates of potential GDP. This finding occurs not just in the United States but across other countries as well and is true for a range of sources of estimates of potential GDP.

We find a similar set of results when we focus on government spending shocks. Regardless of the identification strategy, increases in government spending have transitory effects on GDP, but estimates of potential GDP again display a delayed response to these shocks, ultimately responding to the shock in the same direction as the short-run response of GDP. As with the effects of monetary shocks, the fact that estimates of potential GDP respond so unambiguously to these shocks strongly suggests that real-time estimates of potential GDP are failing to adequately distinguish between permanent and transitory shocks. In this respect, estimates of potential GDP are *sensitive to cyclical fluctuations in GDP* originating from demand shocks.

Turning to supply shocks that should affect potential GDP, the results are more mixed. With productivity shocks, which have immediate and persistent effects on GDP, we find that estimates of potential GDP again respond only very gradually but, after several years, fully incorporate the effects of new productivity levels. With tax shocks, we similarly observe that, after a long delay, estimates of potential GDP eventually catch up to actual changes in GDP. Hence, these two supply shocks provide evidence that real-time estimates of potential output ultimately embody some changes in potential GDP. However, the very slow rate at which information about these shocks is incorporated into estimates of potential GDP points to *an insufficient sensitivity of these estimates in response to supply shocks*. With oil price shocks, however, an even more severe problem arises. We observe persistent declines in GDP after these shocks, but estimates of potential GDP actually go in the opposite direction. As with demand shocks, this specific type of supply shock therefore also presents a challenge to the view that estimates of potential GDP are actually capturing what they are meant to.

Furthermore, we can consistently reproduce the way in which estimates of potential GDP respond to shocks by applying a one-sided Hodrick-Prescott (HP) filter to real-time GDP data. In the U.S. as well as in the cross-country data, this approach generates impulse responses to shocks that are nearly indistinguishable from those found using the actual estimates of potential GDP from all organizations, including the countercyclical behavior of measured potential GDP after oil supply shocks. The HP filter is effectively just a weighted moving average of recent GDP changes, and by construction it does not differentiate between the underlying sources of changes in GDP, be they monetary, technological, or others. Thus, a reliance on simple statistical filters like HP by official agencies could readily rationalize why one might observe a gradual response by real-time measures of potential output to any economic shock, even those that have only transitory effects on GDP and that should presumably be stripped out of estimates of potential GDP.

Fortunately, other approaches to identifying potential output can do better. For example, the approach taken by Olivier Blanchard and Danny Quah (1989) to identify supply and demand shocks can successfully generate real-time estimates of potential output that are consistent with theoretical predictions. Indeed, when Blanchard and Quah's approach is applied to real-time data to recover potential output measured as the historical contribution of shocks with permanent effects on output, the resulting real-time estimate of potential output reacts strongly to identified supply shocks (TFP, taxes, and oil price shocks), and it does not respond significantly to identified demand shocks (monetary policy and government spending shocks). Hence, it does not suffer from the problems associated with most other measures of potential output. Furthermore, this approach yields a starkly different interpretation for changes in U.S. potential output since the Great Recession. Our estimates imply that the gap between potential and actual output in the U.S. increased by about 5 log percentage points between 2007:Q1 (when the gap was likely close to zero) and 2017:Q1, leaving ample room for policymakers to close this gap through demand-side policies, if they chose to do so.

We find similar evidence of a large output gap using other methods to calculate measures of potential output, such as the ones proposed by Jordi Galí (1999), which uses information from labor productivity and hours, or by John Cochrane (1994), which brings in additional information from consumption. Using information from inflation to make inferences about potential output through an estimated Phillips curve also points toward significant slackness. All these methodologies give similar results, pointing to an increase in the gap of 5 to 10 percentage points between 2007:Q1 and 2017:Q1. This assures us that this result is not an artifact of the Blanchard-Quah approach and instead is a feature that is robust to different identification schemes. The idea that significant slackness remained in the U.S. economy through 2017 is also consistent with the low levels of capacity utilization, contained wage growth, and the evolution of labor force participation since the Great Recession.

This paper touches on several bodies of literature. It is most directly tied to recent work since the Great Recession focusing on the possibility of hysteresis—that is, cases where demand shocks lead to permanent effects on the level of economic activity. Though many mechanisms can generate such effects—for example, less research and development during periods of low investment, as shown by Diego Anzoategui and others (2016), Gianluca Benigno and Luca Fornaro (2018), and Patrick Moran and Albert Queralto (2018)—empirical evidence on hysteresis remains scant, as emphasized by Blanchard (2017), with most estimates of monetary and government spending shocks being consistent with the null hypothesis that these shocks have no permanent effects on GDP (for reviews of the literature on monetary and government spending shocks, see Nakamura and Steinsson 2017; Ramey 2016). Recent research has focused on the degree to which the sustained declines in output since the Great Recession have ultimately been interpreted as reflecting declines in potential GDP and therefore can be expected to be long-lasting. Laurence Ball (2014) documents that for most advanced economies, much of the declines in output since the Great Recession have been matched with declines in estimates of potential output. Antonio Fatas and Lawrence Summers (2018) focus on the degree to which fiscal consolidations map first into output changes and then into changes in estimates of potential GDP, with the latter being an indicator that GDP changes will be permanent. *Our results suggest that one should draw little inference from the evolution of estimates of potential GDP about the persistence of GDP changes*; these estimates fail to exclusively identify supply shocks that should drive potential GDP and instead also respond to transitory demand shocks. The fact that most of the output declines observed since the Great Recession are now attributed to declines in potential GDP implies little, other than that these declines have been persistent because estimates of potential GDP fail to adequately distinguish between the underlying sources of changes in GDP.

Our paper also relates to research on news shocks and beliefs about long-run productivity. A strand of the literature studies how news about future productivity can have contemporaneous effects on economic activity long before the productivity changes actually occur (for example, Beaudry and Portier 2006; Barsky and Sims 2011, 2012). In this spirit, Blanchard, Guido Lorenzoni, and Jean-Paul L'Huillier (2017) show that revisions in estimates of future potential output are correlated with contemporaneous changes in consumption and investment. If estimates of future potential output were invariant to transitory shocks, then one could entertain a causal interpretation of these correlations as reflecting the effect of news about the future on current economic decisions. But our results call for caution with this type of interpretation; estimates of potential GDP display sensitivity to demand shocks, and this sensitivity calls into question the basis for causal inference of the type made by Blanchard, Lorenzoni, and L'Huillier (2017).

A third strand of the literature on which we build focuses on the implications of real-time measurement of the output gap for monetary policy. Athanasios Orphanides and Simon van Norden (2002), for example, illustrate how real-time estimates of potential GDP can, in short samples, be sensitive to the method used to measure either the trend or deviations from it. Orphanides (2001, 2003, 2004) argues that the Federal Reserve's mismeasurement of the output gap in the 1970s was one of the primary reasons why inflation was allowed to rise so sharply in the 1970s. We are similarly interested in the difficulties with measuring potential output and the output gap; but rather than studying how sensitive estimates of potential output can be to the different statistical techniques used to identify it, we instead characterize whether the historical estimates of potential output from public and international organizations respond to the "correct" shocks. Our estimates imply that just as the Federal Reserve likely overstimulated the economy in the 1970s because of mismeasurement of potential output, it is now at risk of understimulating the economy by underestimating its productive capacity.

Finally, by comparing actual responses of output after economic shocks to the predictions of agents about these variables, our paper is closely related to recent work studying the expectations formation process of economic agents. Coibion and Gorodnichenko (2012) study the forecast errors of agents to economic shocks and find that these errors are persistent after shocks, consistent with models where agents are not fully informed about the state. By comparing the long-run response of GDP with estimates of potential GDP, this paper similarly provides some insight about how these potential GDP estimates are formed.

The paper is organized as follows. Section I presents information about the estimates of potential output used in the paper. Section II presents our baseline estimates, using U.S. data, of how measures of potential GDP respond to economic shocks. Section III extends these results to a broader range of countries. Section IV presents examples of how estimates of potential output can be improved. And section V concludes.

I. How Estimates of Potential Output Are Created and Used

A seminal description of potential output is in Arthur Okun's (1962) presidential address to the American Statistical Association. Although the notion of potential or natural levels of output had been discussed as far back as research done by Knut Wicksell (1898) and John Maynard Keynes (1936), Okun (1962) provided a sharper definition than had been previously utilized as well as guidance about how to estimate potential output (Hauptmeier and others 2009). Okun emphasized that potential output is a "supply concept, a measure of productive capacity." But it is not designed to represent the maximum amount that an economy could produce. Instead, Okun defines it as the amount that could be produced without generating inflationary pressure. Hence, though potential GDP is related to the nonaccelerating inflation rate of unemployment (NAIRU), potential output provides a more comprehensive assessment of how much an economy can produce without triggering above-normal inflation. This interpretation of potential output advocated by Okun serves as the foundation for most approaches to estimating potential output.

Although Okun proposed to estimate potential output through a combination of knowing the NAIRU and applying what subsequently became known as Okun's law, few organizations follow the specific approach suggested by Okun. As classified by Frederic Mishkin (2007), there are three broad classes of methods to construct a measure of potential output: statistical, production function, and structural (based on dynamic stochastic general equilibrium, DSGE). We first review these methods and then discuss how various agencies measure potential output.

Statistical methods typically impose little theoretical structure on the properties of potential output and interpret low-frequency variation in output series as potential output. One example of this approach is to use univariate time series methods, such as autoregression (AR) models or different types of filters, on actual output to extract a statistical trend component, which is then identified with potential output. Another example is given by methods using several variables—such as output, unemployment, and inflation—to obtain potential output via an unobserved components model and a Phillips curve (Kuttner 1994; Staiger, Stock, and Watson 1997).

In the production function approach, independent estimates of the different inputs that go into the aggregate production function (for example, labor, capital, and multifactor productivity) are plugged into the production function to obtain potential output. Because the objective is to obtain potential output and not actual output, the estimates of the different inputs must correspond to the concept of the maximum (or "normal") amount of each variable that could be used for production without leading to an acceleration of inflation (for example, the labor force participation rate and a level of natural unemployment should be used instead of the cyclical level of employment). In the latter sense, this approach to estimating potential output remains in the spirit suggested by Okun. This approach is also related to growth accounting, because after log-differentiation of a Cobb–Douglas production function, the growth of potential output can be expressed as the weighted average of the growth rates of the different inputs (for an application of this approach to the dynamics of output in the post–Great Recession period, see Fernald and others 2017).

Finally, structural approaches use DSGE models, typically with a New Keynesian structure, to back out potential output. This requires calibrating or estimating the parameters of the model to the relevant economy so that the different shocks hitting the economy can be identified. Once this stage is completed, potential output can be obtained from the solution of the model when certain shocks and frictions are turned off (for example, Andres, Lopez-Salido, and Nelson 2005). This methodology is particularly dependent on models and relies heavily on the estimation of a sophisticated model, which, given limited variation in macroeconomic data, may be a challenge for identification of structural parameters and shocks. Furthermore, because estimated DSGE models have only been used in recent years, no historical, real-time data are available to assess their properties.

The implicit assumptions about the nature of potential output are not identical across methods. The production function approach, for example, explicitly tries to strip out cyclical factors from estimates of potential output. Statistical filters similarly try to separate cyclical fluctuations in output from changes in the trend, with the latter being equivalent to potential. In contrast, with a New Keynesian DSGE model, where the potential level of output reflects counterfactual outcomes under flexible prices, transitory "demand" shocks like temporary changes in government spending can affect the level of potential output for some time, whereas they would be excluded from estimates of potential under the other two approaches (see Blanchard 2017). Because our empirical strategy involves studying the response of real-time estimates of potential output to supply (long-lived) versus demand (transitory) shocks, we are adopting an interpretation of potential output that hews most closely to the production function and statistical filtering approaches, in part because this is precisely the conceptual framework that is most often used by statistical and other agencies when they construct estimates of potential.

I.A. The Congressional Budget Office

The CBO uses the production function approach for estimating potential output. As described by the CBO (2001, 2014), this institution estimates potential output with different methods for five sectors in the economy. The main one is the nonfarm business (NFB) sector, which represents about 75 percent of the U.S. economy. The remaining four smaller sectors are agriculture and forestry, households, nonprofit organizations serving households, and government.

In each of these sectors, the CBO projects the growth of each input by estimating a trend growth rate for it during the previous and current business cycles (as dated by the National Bureau of Economic Research) and by extending that trend into the future. This implies that the trend growth for inputs depends on recent history and on business cycle dating, with possibly large changes in trends when a new business cycle begins. The CBO tries to remove the cyclical component of the growth rate of different variables by estimating the relationships between those variables and a measure of the unemployment rate gap, the difference between the actual unemployment rate and the natural rate of unemployment.

For the NFB, the CBO uses a production function with three inputs: potential labor, services from the stock of capital, and the sector's potential TFP. For the agriculture and forestry sectors, and for nonprofits serving households, potential output is estimated using trends in labor productivity for those sectors. For the household sector, potential output is obtained as a flow of services from the owner-occupied housing stock. Finally, for the government sector, potential output is estimated using trends in labor productivity and depreciation of government capital. The CBO's real-time estimates of potential output have been available since 1991 at an annual frequency and since 1999 at a semiannual frequency.

Estimates of potential output by the CBO play an important role in fiscal policy discussions in the United States. When new tax or spending policies are under review by the U.S. Congress, their implications for future tax revenues, government expenditures, and deficits are assessed under assumptions about the long-run future path of the economy, as captured by estimates of potential GDP (although some policies require the CBO to make inferences about how these policies themselves may change potential output over time, for example, via "dynamic scoring"). How these estimates are formed and how well they separate cyclical from permanent shocks therefore matters for how well these policy measures are scored.

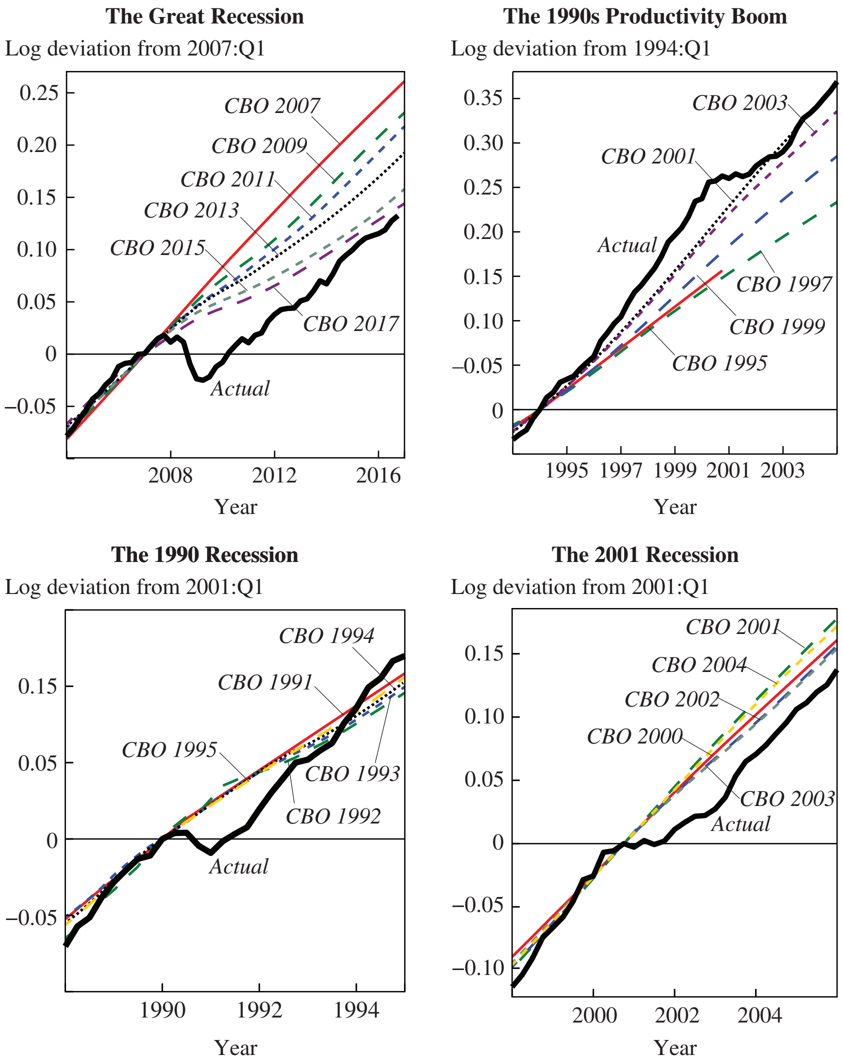
These estimates of potential output are sometimes subject to very large revisions. Preceding the revisions over the course of the Great Recession, for example, the CBO had similarly made a sequence of large *upward* revisions to the projected path of potential output over the course of the 1990s, as illustrated in panel B of figure 1. These upward revisions were tied to the higher-than-expected productivity growth in the U.S. over this period.[1] Other episodes reveal less dramatic sequences of revisions. For example, panels C and D of figure 1 illustrate the CBO's revisions during the two previous U.S. recessions. In both cases, the CBO first started reducing its predicted path of potential output during the recession but then ultimately raised them back up again. In the case of the 1990 recession, GDP ultimately overtook estimates of potential output, whereas over the same time horizon of three years after the start of the recession, the CBO continued to estimate a large output gap after the 2001 recession. But in neither case do we observe a systematic pattern of downward revisions toward the path of actual GDP such as that which was observed after the Great Recession.

I.B. The Federal Reserve

While preparing macroeconomic projections (historically known as Greenbook forecasts) for meetings of the Federal Open Market Committee (FOMC), the staff members of the Federal Reserve Board construct

1. Although it is true that some of these revisions were not related to productivity changes—such as the ones coming from the shift to chained GDP, the addition of software, or revisions to the National Income and Product Accounts—CBO (2001, 2) summarized one of the larger revisions as follows, "CBO also altered its method to address changing economic circumstances. In particular, labor productivity has been growing much faster since 1995 than its post-1973 trend. Because that acceleration has coincided with explosive growth in many areas of information technology (IT), . . . many observers have speculated that the U.S. economy has entered a new era, characterized by more rapid productivity growth. . . . After analyzing the data and the relevant empirical literature, CBO has concluded that elements of the so-called IT revolution . . . explain much of the acceleration in the growth of labor productivity during the late 1990s. CBO has incorporated many of those elements into its economic projections."

Figure 1. Historical Revisions in the CBO's Estimates of U.S. Potential Output[a]



Source: Congressional Budget Office.

a. This figure plots estimates of U.S. potential output from the Congressional Budget Office made at different time periods (that is, at the beginning of the corresponding year). The heavy solid line represents real GDP in the U.S. In each panel, each series is normalized to zero—for 2007, in panel A; for 1994, in panel B; for 1990, in panel C; and for 2000, in panel D.

a measure of the output gap (that is, the difference between actual and potential output) to assist the FOMC's members in their decisionmaking. As pointed out by Rochelle Edge and Jeremy Rudd (2016, 785), from the Board of Governors of the Federal Reserve System, the estimate of the output gap from the Greenbook "is judgmental in the sense that it is not explicitly derived from a single model of the economy. In particular, the staff's estimates of potential GDP pool and judgmentally weight the results from a number of estimation techniques, including statistical filters and more structural model-based procedures."

While describing the evolution of measuring potential output by the Fed, Orphanides (2004, 157) mentions that in the Greenbook estimates, "the underlying model for potential output was a segmented/time-varying trend. The specific construction methods and assumptions varied over time. During the 1960s and until 1976, the starting point was Okun's (1962) analysis. From 1977 onward, the starting point was Clark's (1979) analysis and, later, the related methods explained in Clark (1982) and Braun (1990). Throughout, these estimates of potential output were meant to correspond to a concept of noninflationary 'full employment.' However, judgmental considerations played an important role in defining and updating of potential output estimates throughout this period, so the evolution of these estimates cannot be easily compared to that of estimates based on a fixed statistical methodology."

More recently, Charles Fleischman and John Roberts (2011) describe a methodology to compute potential output using a multivariate unobserved components model that is taken into account by the Federal Reserve Board when producing its judgmental estimates of potential output. Its procedure embeds some parts of many of the methodologies described above; it uses multivariate statistical methods, trend estimation, growth accounting (as in the production function approach), and the relationship between cyclical fluctuations in output and unemployment (as in Okun's law). The authors use data on nine macroeconomic series: real GDP; real gross domestic income; the unemployment rate; the labor force participation rate; aggregate hours for the NFB; a measure of NFB sector employment; two measures of NFB sector output (measured on the product side and on the income side); and inflation as measured by the Consumer Price Index, excluding food and energy. The common cyclical component of the economy is constrained to follow an AR(2) process, and trends in the series are related to each other via structural equations (for example, Okun's law, production function) to obtain a final measure of the trend of output, which is associated with potential output.

Real-time estimates of potential output can be computed from the estimates of actual output and the output gap reported in Greenbooks since 1987.[2] Real-time estimates for the same variables in the 1969–87 period are provided by Orphanides (2004). For this earlier period, the quality of the estimates is likely to be worse because the estimates sometimes had to be obtained from a variety of sources (for example, the Council of Economic Advisers) other than the Federal Reserve. As a result, we take the 1987–2011 series as the benchmark and explore the longer time series in robustness checks. Because the Greenbooks only forecast potential output growth for up to a few years, we cannot reproduce figure 1 (the evolution of real-time forecasts of potential GDP during the Great Recession) for Greenbook forecasts.

Estimates of potential output play an immediate role in decisionmaking by the Federal Reserve. One of the objectives of the FOMC is to stabilize output around potential, and whether output is below or above potential is commonly interpreted as having implications for inflation, the other objective targeted by the Federal Reserve. Potential mismeasurement of the output gap (the difference between actual output and potential) is mentioned (for example, Orphanides 2001) as a reason why the Federal Reserve allowed inflation to rise during the 1970s, and Fed chairman Alan Greenspan's perception that potential output was growing unusually rapidly in the 1990s explains why, over this period, monetary policymakers were less concerned about inflation than they normally would have been, given the low unemployment rates (Gorodnichenko and Shapiro 2007).

I.C. The International Monetary Fund

The IMF provides estimates of potential output for a wide range of countries. There is considerable methodological variation across countries in how the IMF generates estimates of potential output. As summarized by Carlos de Resende (2014, 24), in a study conducted by the IMF's Independent Evaluation Office, "Interviews with staff showed that the use of the macro framework is country-specific and varies greatly in detail and sophistication, ranging from the use of 'satellite' models to simply entering numbers based on judgment." In this respect, the IMF's approach to measuring

2. This series is available from the Real-Time Data Research Center at the Federal Reserve Bank of Philadelphia. There is a five-year delay period for the release of Greenbook projections.

potential output is methodologically similar to measures reported in the Greenbooks, in the sense that they use a combination of different methods to compute potential output and then aggregate them using a great deal of judgment. At the same time, the IMF staff often uses the Hodrick–Prescott filter and/or multivariate methods such as the ones described in Patrick Blagrave and others (2015) to construct measures of potential output. The IMF provides potential output estimates for 27 countries.[3] Nowcasts and one-year-ahead forecasts are available for the period 2003–16. Since 2009, the IMF has also provided up to five-year-ahead forecasts for potential output.

Estimates of potential output can play an important role in the IMF's policy decisions. To assess the sustainability of countries' fiscal policies, tax and spending levels are commonly evaluated at the level of potential GDP to control for the cyclical changes in revenues and expenditures that are expected to be transitory, thereby helping to gauge any "structural" fiscal imbalances. These imbalances are then the primary focus of policy reforms undertaken by those countries receiving funds from the IMF during times of crisis.

I.D. The Organization for Economic Cooperation and Development

The OECD's estimates of potential output are based on a production function approach. In particular, the OECD uses a Cobb–Douglas production function with constant returns to scale that combines physical capital, human capital, labor, and labor-augmenting technological progress. Each of these inputs is projected using a trend, and TFP is assumed to converge to a certain degree among different countries in the medium run. As pointed out by the OECD (2012, 195): "The degree of convergence in total factor productivity depends on the starting point, with countries farther away from the technology frontier converging faster, but it also depends on the country's own structural conditions and policies." Note that when forecasting potential output in the medium term,

3. These countries are Australia, Austria, Belgium, Canada, Switzerland, Cyprus, the Czech Republic, Germany, Denmark, Spain, Estonia, Finland, France, the United Kingdom, Greece, Hungary, Ireland, Iceland, Italy, Japan, South Korea, Luxembourg, Malta, the Netherlands, Norway, New Zealand, Poland, Portugal, Slovak Republic, Slovenia, Sweden, Turkey, and the United States. For more information on the time periods for the data on these countries, see online appendix table 1. The online appendixes for this and all other papers in this volume may be found at the *Brookings Papers* web page, www.brookings.edu/bpea, under "Past BPEA Editions."

the OECD assumes that output gaps close over a period of 4 to 5 years, depending on their initial size. Therefore, one should expect to see above average future growth for countries with large output gaps. Relative to the IMF, the OECD covers more countries and has longer time series (see the online appendix). For many countries, nowcasts and one-year-ahead forecasts have been available since 1989. Since 2005, the OECD has also reported five-year-ahead forecasts for potential output. As with the IMF, estimates of potential output in the OECD are commonly used to assess cyclically adjusted fiscal balances and to characterize the need for structural reforms.

I.E. Consensus Economics

Consensus Economics, a global survey firm of professional forecasters, does not provide estimates of potential output, but it does report forecasts for the growth rate of actual output from 1 to 10 years into the future. Because estimates made for several years into the future (for example, years 6 through 10) are likely to be independent of business cycle conditions, we use these long-run estimates as an approximation of the growth rate of potential output at the same horizon. These data are available for 12 countries, and the starting date varies across countries from 1989 to 1998 (see online appendix table 1). Given the wide range of forecasters included in the Consensus Economics forecasts, one cannot readily summarize how these forecasts are made. Private forecasts, however, are widely used in both public and international organizations for comparison purposes with in-house forecasts.

I.F. Comparison of Potential Output Measures

Table 1 documents some basic moments for estimates of the potential output growth rate (nowcasts) produced by the IMF and OECD, as well as the forecasted long-term output growth rate from Consensus Economics. We work with growth rates of potential output rather than levels because the definition of output varies across time (base year) and agencies. The growth rate series are highly correlated and generally have similar moments across sources. This is especially true for the IMF and OECD forecasts, which conceptually are measuring the same objects (nowcasts of potential GDP). The Consensus Economics forecasts, in contrast, are at a different horizon and are for actual rather than potential GDP. These strong correlations are not driven by outliers. Indeed, there are few large differences across sources, and these tend to be concentrated in a handful of countries and periods (see online appendix figure 1).

Table 1. Comparison Output Measures from the IMF, OECD, and Consensus Economics[a]

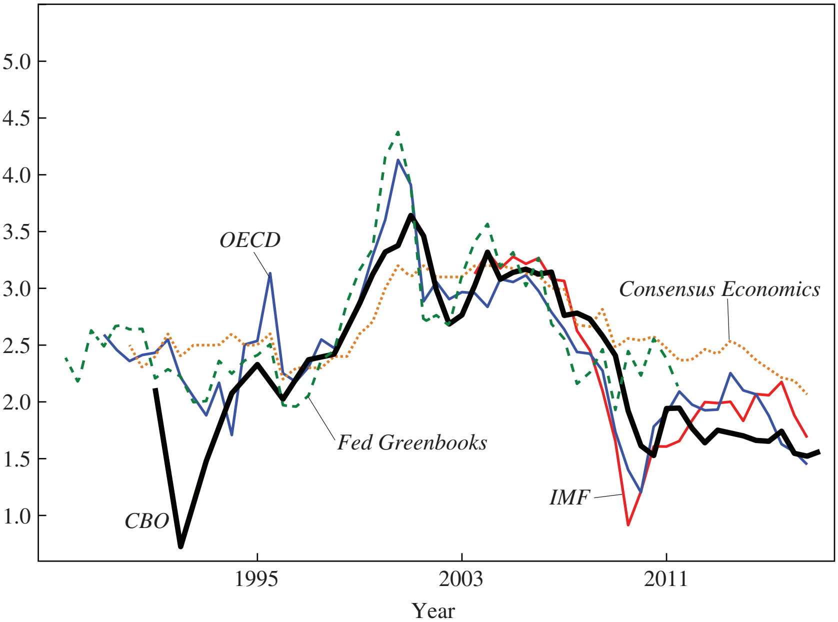
	Institution and output measure		
Basis for comparison and correlation	*IMF: Potential output growth rate (nowcast)*	*OECD: Potential output growth rate (nowcast)*	*Consensus Economics: 6- to 10-year-ahead forecast for actual output growth rates*
Observations	607	1,358	581
Mean	1.64	2.30	2.22
Standard deviation	1.10	1.25	0.54
Correlation			
IMF	1.00		
OECD	0.87	1.00	
Consensus Economics	0.72	0.78	1.00

a. This table reports moments of measures of potential output from the IMF and OECD across the countries described in online appendix table 1 and listed in footnote 3 in the text, as well as moments of forecasted growth rates of GDP 6 to 10 years ahead from Consensus Economics. See subsection I.F for details.

Figure 2 illustrates that this strong correlation across series is not restricted to differences in growth rates across countries. Time series for the growth rate of U.S. potential output across the different institutions that produce estimates (Fed Greenbooks, CBO, IMF, OECD, and Consensus Economics long-term forecasts of actual output) track each other closely as well. There are nonetheless occasional differences across estimates. After the 1990–91 recession, for example, the CBO reduced its estimate of potential GDP growth significantly more than the staff of the Federal Reserve Board, whereas private forecasters hardly changed their long-term forecasts of growth at all. After the Great Recession, the IMF and OECD both lowered their estimates of potential GDP growth far more than the Fed Greenbooks or the CBO, but then revised them back up while the CBO continued to progressively revise its estimates of potential GDP growth down.

Figure 3 plots a longer time series of estimates of potential GDP available from the Fed Greenbooks, as extended backward by Orphanides (2004). In addition, we plot several statistical approaches to estimating potential GDP, including a one-sided, five-year moving average of real-time GDP and a one-sided HP-filter ($\lambda = 500{,}000$) of real-time GDP. The HP filter tracks the Greenbooks' estimates of potential output quite closely, especially since the mid-1980s, while the moving-average approach tends to display larger fluctuations. All series co-move relatively closely with a moving average of capacity-adjusted TFP changes as measured by John Fernald (2012).

Figure 2. Estimates of Potential U.S. Output Growth Rate and Forecasted Long-Term Growth for Actual Output[a]



Sources: International Monetary Fund; Congressional Budget Office; Organization for Economic Cooperation and Development; Federal Reserve Greenbooks; Consensus Economics.

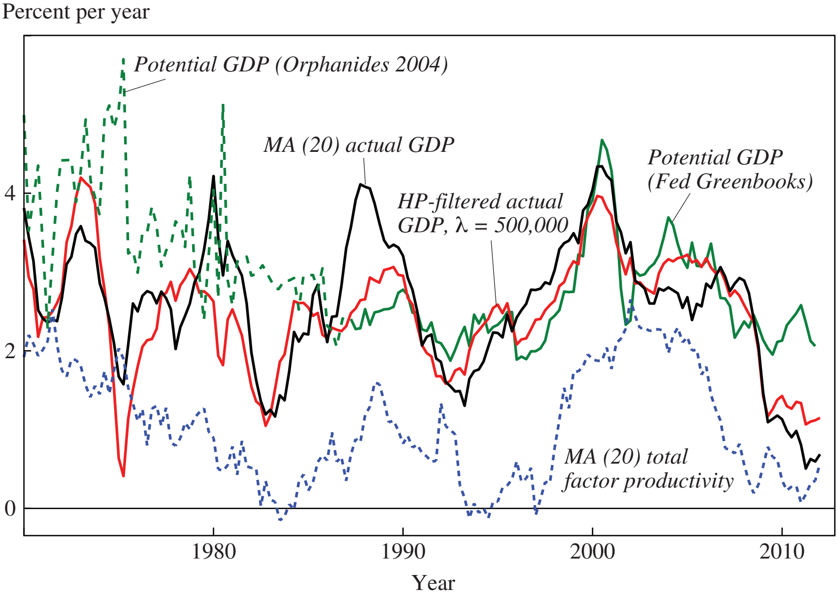
a. All series in the figure are real-time data at the semiannual frequency. The potential outputs for the IMF, OECD, and CBO are reported for the current calendar year. Potential output for Greenbooks is the semiannual average of quarterly growth rates of potential output for the quarters in a given semester. Series for Consensus Economics show the 6- to 10-year-ahead forecast for the actual output growth rate (per year).

The persistence in revisions of potential GDP shown in figures 2 and 3 suggests that some of these revisions might be predictable from recent changes. We evaluate this formally by regressing revisions of potential GDP on lags of itself:

$$(1) \quad \left(\Delta \log Y^*_{t|t} - \Delta \log Y^*_{t|t-1}\right) = \alpha + \beta\left(\Delta \log Y^*_{t-1|t-1} - \Delta \log Y^*_{t-1|t-2}\right) + error_t$$

where $\Delta \log Y^*_{t|t}$ is the growth rate of potential output in time t according to a projection made at time s. We find (table 2) a mild amount of predictability in the Greenbooks' revisions of potential GDP. With the CBO, the coefficient on lagged revisions is similar but not significantly

Figure 3. Real-Time Estimates of U.S. Potential Output Growth Rate and Trends in Actual Output Growth Rate[a]



Sources: Orphanides (2004); Federal Reserve Greenbooks; Federal Reserve Bank of Philadelphia.

a. All series are real time at the quarterly frequency. Potential output for the pre-1987 period is from Orphanides (2004). Potential output for 1987–2011 is from the Federal Reserve Bank of Philadelphia. Potential output is measured as the growth rate of potential output between a given quarter and the next three quarters. *HP-filtered actual output* (HP = Hodrick–Prescott) is calculated as the value of the one-sided HP filter trend for the quarter given the first vintage of GDP data that covers the given quarter, with HP filter smoothing parameter of 500,000. *MA (20) actual output* is calculated as the 20-quarter moving average over the current and preceding 19 quarters reported in the first vintage of GDP data that covers the given quarter. *MA (20) total factor productivity* for a given quarter is calculated as the 20-quarter moving average running on the current quarter and the preceding 19 quarters.

different from zero. The results are different for international data, with coefficients on past OECD revisions being not different from zero and with those on past IMF and Consensus Economics revisions exhibiting negative predictability.

II. How Estimates of U.S. Potential Output Are Adjusted after Economic Shocks

Although a limited unconditional predictability is a desirable attribute of estimates of potential GDP, it does not imply that there is no predictability in estimates of potential output *conditional* on different economic shocks.

Table 2. Predictability of Revisions in Estimates of Potential GDP[a]

	Source				
Dependent variable: ($\log Y^*_{t\|t} - Y^*_{t\|t-1}$)	*CBO* *(1)*	*Fed Greenbooks* *(2)*	*OECD* *(3)*	*IMF* *(4)*	*Consensus Economics* *(5)*
($\log Y^*_{t-1\|t-1} - \log Y^*_{t-1\|t-2}$)	0.204	0.294***	–0.066	–0.154***	–0.355***
	(0.132)	(0.086)	(0.040)	(0.044)	(0.045)
Observations	42	96	1,282	548	566
R^2	0.065	0.085	0.163	0.351	0.288
Number of countries			31	27	12

a. This table presents regressions of the revision in estimates of potential GDP on the previous revisions in estimate of potential GDP (equation 1). Newey–West standard errors are in parentheses. "Source" indicates where estimates of potential output come from the Congressional Budget Office (CBO), Greenbooks of the Federal Reserve Board, the Organization for Economic Cooperation and Development (OECD), the International Monetary Fund (IMF), and Consensus Economics. For Consensus Economics, revisions are for the growth rate of GDP at horizons of 6 to 10 years. Columns 3–5 are across countries and include time and country fixed effects. Within R^2 is reported for columns 3–5.

To assess how estimates of potential output respond to economic shocks, we combine the estimates described in the previous section with identified measures of economic or policy shocks.

II.A. Measures of Economic Shocks

There is an extensive literature on identifying shocks that potentially drive business cycle and longer-term fluctuations, particularly for the United States (for a survey, see Ramey 2016). Following this literature, we employ several measures of both "demand" and "supply" shocks for the U.S. Our use of the terms "supply" and "demand" reflects a certain abuse of terminology. All the shocks we consider have both supply and demand effects in modern business cycle models. Our classification instead primarily relies on whether these shocks appear to have permanent or transitory effects on GDP. We define demand shocks as those whose real effects appear to be transitory and therefore should not affect estimates of potential output.[4]

For supply shocks, we consider changes in TFP, oil price shocks, and tax shocks. TFP changes are measured as by Fernald (2012), who adjusts Solow residuals for time-varying utilization of inputs. Although these data are somewhat sensitive to vintage (see Kurmann and Sims 2017), we rely on the final vintage of the data because the data by vintage are available for relatively recent times. For oil price shocks, we use oil supply shocks as

4. Because the units of these shocks vary, we normalize all shocks to be mean zero and have unit variance.

identified by Lutz Kilian (2009).[5] For tax shocks, we use Christina Romer and David Romer's (2010) narrative measure of exogenous tax changes. To be clear, tax shocks have both demand and supply effects. We denote them here as "supply" shocks because Romer and Romer (2010) document that they have permanent effects on output, and therefore should be captured by estimates of potential GDP.

We consider three identified demand shocks, all related to policy. The first are monetary policy shocks. For the United States, our baseline measure of these shocks follows the quasi-narrative approach of Romer and Romer (2004). They use the narrative record to construct a consistent measure of policy changes at FOMC meetings since 1969, then orthogonalize these policy decisions to the information available to policymakers at each FOMC meeting, as captured by the Greenbook forecasts prepared by the staff of the Federal Reserve Board before each FOMC meeting. The unexplained policy changes are then defined as the monetary shocks. We use the updated version of these shocks from Coibion and others (2017) and set values after the onset of the zero lower bound equal to zero.[6]

The second type of demand shock we consider are the military spending news shocks given by Valerie Ramey (2016). Using real-time measures of the expected future path of defense spending in the United States, Ramey constructs a measure of the present discounted value of future defense expenditures for each quarter. Changes in these measures from one quarter to the next thus reflect changes in either current or future defense spending.

Finally, we consider a broader measure of government spending shocks, namely, differences between ex-post government spending and ex-ante forecasts of this spending following Alan Auerbach and Gorodnichenko (2012b). Unlike the Ramey news measure, this measure captures unanticipated short-run changes in government spending but is broader in that it includes more than just military spending.

All three types of demand shocks have repeatedly been found to have only transitory effects on GDP (see Nakamura and Steinsson 2017; Ramey 2016), so there is little evidence supporting the hysteresis hypothesis that transitory shocks have long-lived effects on output (and therefore potential) through endogenous productivity or tax responses. As emphasized by

5 We also tried using the oil shocks identified by Baumeister and Hamilton (2015) in place of the ones identified by Kilian (2009). The results were very similar and are available from the authors upon request.

6. We also experimented with monetary policy shocks identified via recursive ordering of vector autoregression residuals, as done by Bernanke and Blinder (1992), and we found similar results, as documented in online appendix figure 5.

Blanchard (2017), these transitory shocks could still affect potential GDP in a transitory fashion in the presence of physical or human capital. As a result, we study not just the response of nowcasts of potential GDP to these shocks but also of long-run forecasts of potential from the CBO as well as long-run forecasts of GDP growth from private forecasters. The latter two should unambiguously not respond to these transitory shocks. Finally, even if the real world were characterized by hysteresis, monetary policymakers explicitly rule out this channel and emphasize that, in their view, monetary policy has only transitory effects on GDP.[7] Their estimates of potential GDP should therefore be invariant to monetary shocks.

II.B. Effects of Shocks on Actual Output and Estimates of Potential Output in the United States

To provide a benchmark for how we might expect estimates of potential output to respond to economic shocks, we first characterize the response of actual output to these shocks. Specifically, we regress ex-post changes in output on current and past values of a shock, as follows:

$$\Delta \log Y_t = \alpha + \sum_{k=0}^{K} \phi_k \epsilon_{t-k} + error_t, \tag{2}$$

where t indexes time (quarters), $\Delta \log Y_t$ is the growth rate of real GDP, ϵ is an identified shock, and *error* is the residual. A key advantage of this moving-average specification is that it allows us to handle data with mixed frequencies and gaps in the time series as well as correlations of the error term. For consistency, we run these regressions at the same time frequency as what is available for estimates of potential output, namely, quarterly when comparing with Greenbook forecasts, and semiannually otherwise. Because Greenbook forecasts of potential output begin in 1987, we run the regression for output over the same time sample. Given the limited number of observations available, we include only one shock at a time (the shocks are roughly uncorrelated). Because the error term is not necessarily white noise, we use Newey–West standard errors everywhere.[8] Impulse responses

7. For example, in a speech on March 3, 2017, Janet Yellen stated that "monetary policy cannot, for instance, generate technological breakthroughs or affect demographic factors that would boost real GDP growth over the longer run or address the root causes of income inequality. And monetary policy cannot improve the productivity of American workers. Fiscal and regulatory policies—which are of course the responsibility of the Administration and the Congress—are best suited to address such adverse structural trends" (Yellen 2017).

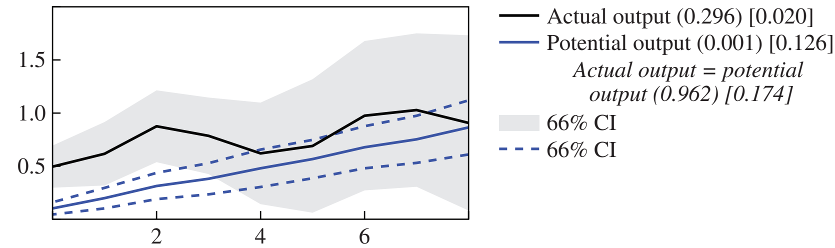
8. Because the null hypothesis we are testing is that of zero response of output and potential output, the fact that shocks are estimated does not constitute an issue for standard errors and tests of the null hypothesis, as shown by Pagan (1984).

Figure 4. Responses of U.S. Output and Greenbook Estimates of Potential U.S. Output to Shocks[a]

Panel A: Supply Shocks

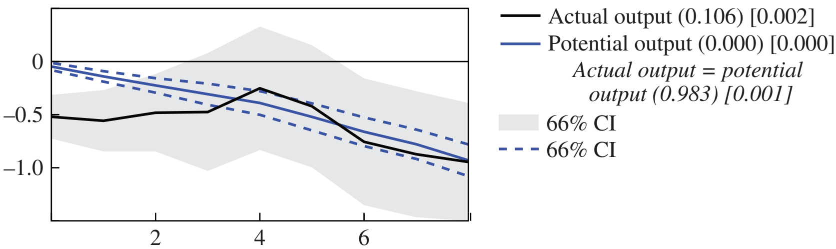
Total factor productivity shock (Fernald 2012)

Output growth rate, percent, annualized



Tax shock (Romer and Romer 2010)

Output growth rate, percent, annualized



Oil supply shock (Kilian 2009)

Output growth rate, percent, annualized

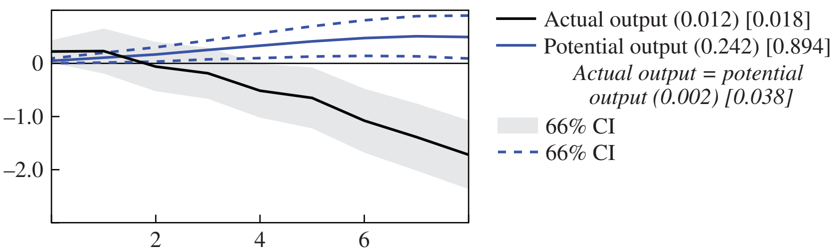
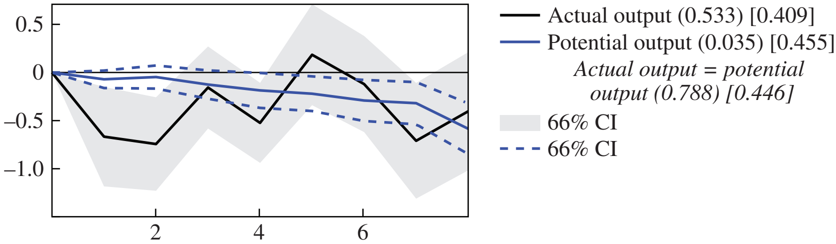


Figure 4. Responses of U.S. Output and Greenbook Estimates of Potential U.S. Output to Shocks[a] (*Continued*)

Panel B: Demand Shocks

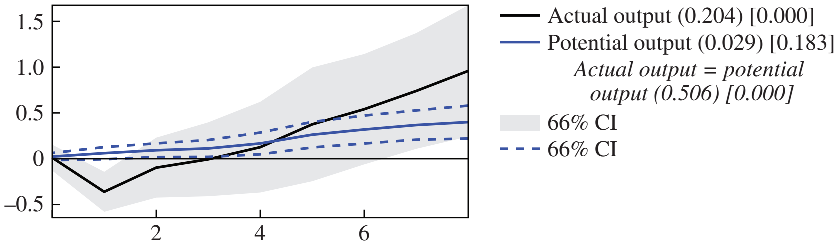
Monetary policy shock (Romer and Romer 2004)

Output growth rate, percent, annualized



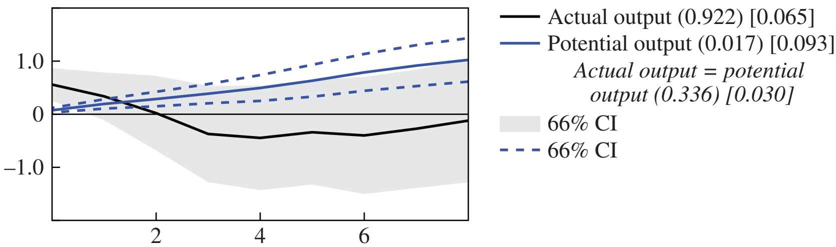
Military spending shock (Ramey 2016)

Output growth rate, percent, annualized



Government spending shock (Auerbach and Gorodnichenko 2012a, 2012b)

Output growth rate, percent, annualized



Sources: Authors' calculations, with potential output from Federal Reserve Greenbooks and identified shocks from Fernald (2012); Romer and Romer (2004, 2010); Kilian (2009); Ramey (2016); Auerbach and Gorodnichenko (2012a, 2012b).

a. This figure reports impulse response functions (IRFs) estimated using equations 2 and 3. The estimation sample covers the longest possible period with nonmissing observations for shocks and potential output (output gap) available at the Federal Reserve Bank of Philadelphia. In parentheses, we report the *p* values for a test of whether the response of actual (potential) output is different from zero at the maximum horizon (eight quarters). In square brackets, we show the *p* values for a test of whether the path of the response of actual (potential) output is different from zero over the entire duration of the IRF. The last row of the legend—for which there is no line in the graphs—reports *p* values for a test of equality of responses of actual and potential output at the maximum horizon (parentheses) and a test of equality of the paths of the responses for actual and potential output are equal across horizons. CI = confidence interval.

come directly from the estimates of ϕ. To recover responses of the *level* of output, we cumulate ϕ_k up to a given horizon. For example, the level responses are ϕ_0 for $h = 0$, $\phi_0 + \phi_1$ for $h = 1$, $\phi_0 + \phi_1 + \phi_2$ for $h = 2$, and so on.[9]

For each impulse response, we include 66 percent confidence intervals and the legend of each associated graph reports the p values for two types of tests. In parentheses we report the p value for a test of whether the response of actual output is different from zero at the maximum horizon (eight quarters), while in square brackets we show the p value for a test of whether the path of the response of actual output is different from zero over the entire horizon of the impulse response. These p values are also included in panel A of online appendix table 2, together with more information that we describe later in the paper.

We plot the responses of actual output to each type of shock in figure 4, which appears on the previous two pages. Panel A of the figure focuses on the three supply shocks. In response to a TFP shock, output immediately rises about 0.5 percentage point and remains persistently higher by about this magnitude. Hence, these TFP shocks appear to have permanent effects on output. Tax increases have a (negative) contemporaneous effect on output that is similarly sustained over the entire impulse response horizon. In contrast, negative oil supply shocks have a more delayed effect on output, but are associated with a long-lived decline in GDP. In short, all three supply shocks have the expected long-lived effects on GDP. As a result, we would expect them to be captured by high-quality measures of potential GDP.

Turning to demand-side shocks (panel B of figure 4), we again find the expected responses of output. Contractionary monetary policy shocks push output down. The point estimates are much less precise than those of Romer and Romer (2004), reflecting the shorter time sample, the fact that monetary shocks are smaller over this limited sample, and the different approach to estimating impulse responses. Increases in expected military expenditures have a delayed positive effect on GDP (which reflects the fact that the expenditures themselves are also generally delayed).[10] Immediate spending shocks, as given by Auerbach and Gorodnichenko (2012b), have transitory, short-run effects on GDP and no long-run effects. Demand-side shocks therefore generally deliver cyclical variation in output but no long-run effects on GDP. As a result, we would expect high-quality measures of potential GDP to be insensitive to these shocks.

9. For monetary policy shocks, we constrain $\phi_0 = 0$ to capture the minimum delay restriction.

10. Although our horizon of impulse responses is too short to illustrate this, Ramey (2016) shows that news about future military spending has only transitory effects on GDP.

To characterize the effects of these economic shocks on estimates of potential output, we run equivalent specifications:

$$\Delta \log Y^*_{t|t} = \alpha + \sum_{k=0}^{K} \phi_k \epsilon_{t-k} + error_t, \tag{3}$$

where $\Delta \log Y^*_{t|t}$ is the (nowcast) estimated growth in potential in quarter t given information in quarter t at an annualized rate. We first consider Greenbook estimates of potential output and extend our results to alternative estimates of potential in subsequent sections. Responses of the implied *level* of potential output are constructed in the same way as before. For comparison, we plot the responses of potential output in the same graphs as the responses of actual output, and we also include 66 percent confidence intervals and the p values for the same tests mentioned above (now for the responses of potential output instead of actual output). Finally, we also include the p values for a test of whether paths of the responses for actual and potential output are equal over the entire duration of the impulse response (in square brackets) and the p values of a test of whether the responses are equal at the maximum horizon (in parentheses). The p values are also included in panel A of online appendix table 2.

Looking first at TFP shocks, we find that estimates of potential GDP respond very gradually but in the same direction as actual GDP. The shock has little immediate impact on estimates of potential; but after two years, the responses are overlapping and estimates of potential GDP have caught up to actual GDP. Very similar results are obtained with tax shocks: Estimates of potential GDP are unchanged immediately after the shock, but gradually converge to the path of actual GDP. Hence, with both TFP and tax shocks, one would ultimately attribute the decline in output to a decline in potential output, but only with some delay. One possible reason for delayed responses of forecasts is information rigidity, as suggested by Coibion and Gorodnichenko (2012, 2015a). However, the fact that estimates of potential GDP evolve very gradually after tax shocks (which occur only for large legislative tax changes of which staff members at the Fed would be well aware) suggests that other mechanisms must be at play to explain the inertia in real-time estimates of potential output.

Turning to the response to oil price shocks, we find a starkly different response: Estimates of potential GDP *increase* over time while actual GDP *falls*. In contrast to TFP and tax shocks, in which the long-run response of output is ultimately matched by the response of potential, contractionary oil price shocks are associated with sharply falling measured output gaps (Y_t/Y^*_t) in the long run, as estimates of potential are progressively increased while output itself is falling. Policymakers facing a trade-off between stabilizing inflation (which rises after a negative oil supply shock, thereby calling for higher interest rates) and closing the output gap (which is falling, calling for lower interest rates) are therefore perceiving an even starker trade-off because the rise in the estimate of potential output makes the output gap seem even more negative.[11] This result is not driven by the specific measure of oil supply shocks (we find a similar result with Kilian's 2008 measure of OPEC supply shocks) or by the sample period (we find similar results for alternative periods).

There are several potential explanations for this finding. One is that policymakers are confounding oil supply and demand shocks: If they observe a supply-driven increase in oil prices that they incorrectly attribute to stronger global demand for oil from, for example, improved technology, then this might lead them to revise their estimates of potential GDP upward, even as actual GDP is falling. An alternative explanation is that higher oil prices might be perceived as inducing greater investment in new energy sources and alternative energy technologies, which could then raise potential GDP in the long run, even as short-run GDP falls, though there is little evidence that GDP ultimately responds in a positive manner. The available data unfortunately do not enable us to identify the underlying explanation. If nothing else, this result provides a surprising example of how estimates of potential GDP can move in the direction opposite to that of actual GDP.

Turning to demand shocks, we again observe important deviations from what one would expect of estimates of potential GDP. With monetary and both types of fiscal shocks, estimates of potential respond little on impact to these shocks but progressively respond in the same manner as the short-run response of GDP. The transitory decline in GDP after a contractionary monetary shock is followed by a persistent decline in the real-time estimates of potential GDP, while the transitory increase in output after an increase in government spending is followed by a persistent rise in estimates of potential GDP. Hence, these *cyclical* fluctuations in output lead to the perception among forecasters that they are *permanently* affecting output, as if they were TFP or tax shocks, despite the fact that their effects on income are actually short-lived.

11. The pronounced decline in the p*erceived* output gap after oil supply shocks is consistent with the view that monetary policymakers were too willing to accommodate these shocks with lower interest rates, and that this accommodation may have exacerbated the Great Inflation of the 1970s.

Our results are not limited to these specific examples of identified shocks. For example, we can identify supply and demand shocks jointly, as was done by Blanchard and Quah (1989), by running a vector autoregression (VAR) with output growth and unemployment and restricting demand shocks to have no long-run effects on output. When we use these supply and demand shocks to characterize the response of real-time estimates of potential output over the same period, we again find that real-time estimates respond very gradually to both shocks, moving in the direction of the change in output (online appendix figure 2). Importantly, because this identification explicitly imposes the fact that only supply shocks have permanent effects on GDP, it addresses the possibility that some demand shocks might have hysteresis effects and therefore should be incorporated into estimates of potential GDP. In short, across identification schemes, we find an overresponse of real-time estimates of potential GDP to demand shocks and an underresponse to supply shocks.

II.C. The Robustness of Baseline Results for the United States

Because of the relatively short samples involved, we want to verify that our results are robust to a range of reasonable variations. Our first check is on the empirical method used to estimate impulse responses. As an alternative to equations 2 and 3, we reproduce impulse responses of actual output and nowcasts of potential GDP to each of the shocks using autoregressive distributed lag specifications to estimate impulse response functions (IRFs), as done by Romer and Romer (2004), namely:

$$(4) \qquad \Delta \log Y_t = \alpha + \sum\nolimits_{j=1}^{J} \delta_j \Delta \log Y_{t-j} + \sum\nolimits_{k=0}^{K} \phi_k \epsilon_{t-k} + error_t,$$

using $J = 4$ and $K = 8$. The results are presented in online appendix figure 3. By and large, the results are very similar. With productivity and tax shocks, we continue to find persistent but delayed effects on estimates of potential GDP that are ultimately converging to the responses of actual GDP. Similarly, with all three demand shocks, we find the same qualitative patterns as with the previous empirical specification. The only difference lies in the response to oil supply shocks, where we no longer observe a pronounced rise in estimates of potential GDP. Instead, our estimates indicate no response of the nowcasts of potential, suggesting some sensitivity in this result.

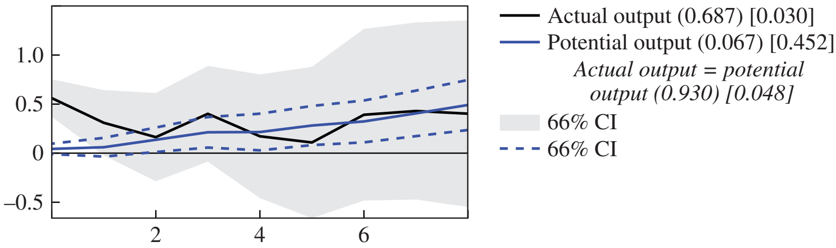
One potential source for this empirical sensitivity is the limited time sample. As a result, we replicate our baseline results over an extended time period, where for each shock we now use the maximum time sample

Figure 5. Responses of U.S. Output and Greenbook Estimates of Potential U.S. Output to Shocks: Extended Sample[a]

Panel A: Supply Shocks

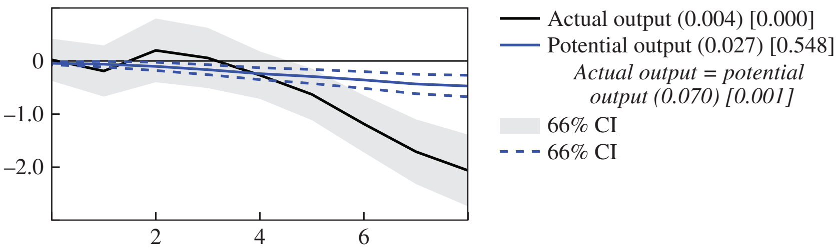
Total factor productivity shock (Fernald 2012)

Output growth rate, percent, annualized



Tax shock (Romer and Romer 2010)

Output growth rate, percent, annualized



Oil supply shock (Kilian 2009)

Output growth rate, percent, annualized

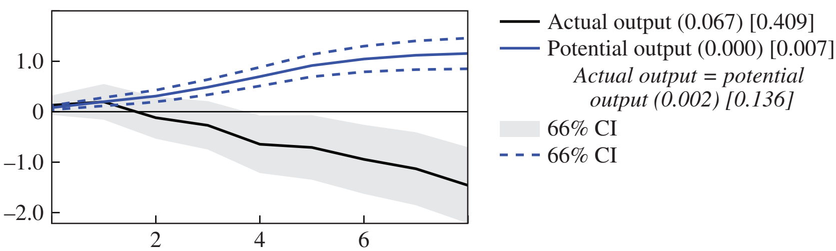
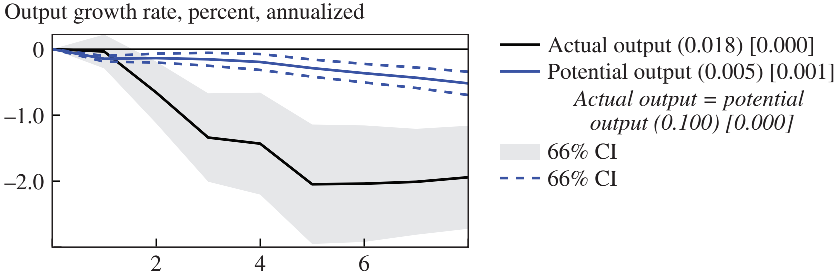


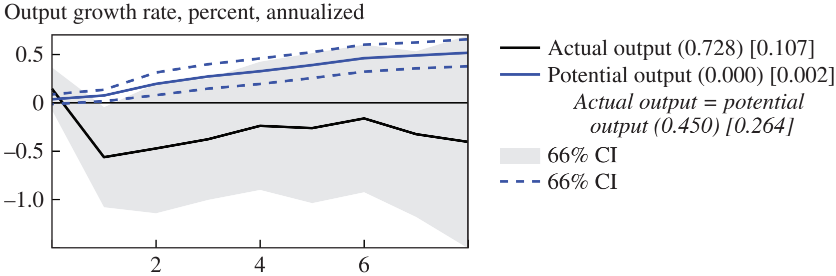
Figure 5. Responses of U.S. Output and Greenbook Estimates of Potential U.S. Output to Shocks: Extended Sample[a] (*Continued*)

Panel B: Demand Shocks

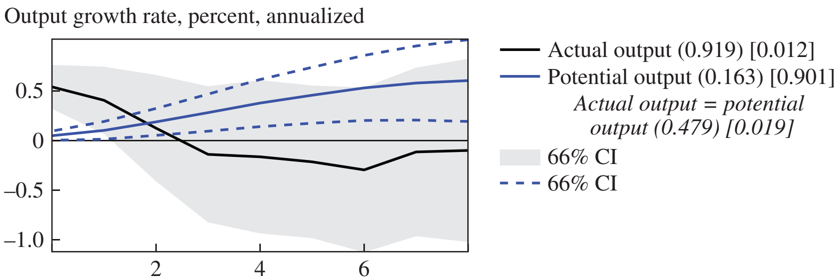
Monetary policy shock (Romer and Romer 2004)



Military spending shock (Ramey 2016)



Government spending shock (Auerbach and Gorodnichenko 2012a, 2012b)



Sources: Authors' calculations, with potential output from Federal Reserve Greenbooks and identified shocks from Fernald (2012); Romer and Romer (2004, 2010); Kilian (2009); Ramey (2016); Auerbach and Gorodnichenko (2012a, 2012b).

a. This figure reports impulse response functions (IRFs) estimated using equations 2 and 3. The estimation sample covers the longest possible period with nonmissing observations for shocks and potential output (output gap) using output gap data starting in 1970. In parentheses, we report the *p* values for a test of whether the response of actual (potential) output is different from zero at the maximum horizon (eight quarters). In square brackets, we show the *p* values for a test of whether the path of the response of actual (potential) output is different from zero over the entire duration of the IRF. The last row of the legend—for which there is no line in the graphs—reports *p* values for a test of equality of responses of actual and potential output at the maximum horizon (parentheses) and a test of equality of the paths of the responses for actual and potential output are equal across horizons. CI = confidence interval.

available across both the shocks and the Greenbook estimates of potential GDP (1969–2011). The results, presented in figure 5 (which appears on the previous two pages), confirm our baseline findings: There is a delayed but persistent response of the estimates of potential GDP to all shocks. In every case but oil supply shocks, the nowcasts evolve in the direction of the short-run changes in GDP. With oil supply shocks, the estimates of potential GDP rise in an even more pronounced fashion, while actual output falls.[12] Hence, the baseline results are not specific to the period since 1987.

We also consider whether our results are sensitive to relying on nowcasts of potential GDP growth. Because the Fed Greenbooks also include forecasts and backcasts of potential GDP growth (two years in each direction), we can characterize how the perceived *path* of potential GDP evolves after each shock. We find very little difference relative to nowcasts, implying that Federal Reserve staff members raise or lower the entire path of projected and past potential GDP growth in response to shocks (online appendix figure 4).

Another potential issue with these results is our reliance on estimates of potential GDP from a single source: the staff of the Federal Reserve Board. In figure 6 (which appears two pages down from here), we reproduce our results using estimates of potential GDP from the Congressional Budget Office. One advantage of CBO estimates is they are available at longer horizons. As a result, we consider both "nowcasts" of potential GDP (equivalent to Greenbook estimates) as well as five-year-ahead forecasts (that is, the growth rate of potential output in five years from the date when a forecast is made). A disadvantage of CBO estimates, as discussed in subsection I.A, is that the sample for these is more limited and the time frequency at which forecasts are available is reduced. Not surprisingly, the effects of each shock on GDP are therefore considerably less precisely estimated. However, the responses of the estimates of potential GDP are still quite precise. Qualitatively, we find that the CBO's estimates of current potential GDP respond much like those from the Greenbooks: gradually but persistently to all shocks. Long-run forecasts of potential GDP generally respond by less than those of current potential GDP. However, they still ultimately respond to demand shocks, implying that the CBO implicitly interprets cyclical shocks as having permanent effects on GDP.

12. When we apply the autoregressive distributed lag specification to oil supply shocks over the whole sample, we find the same result.

The fact that CBO forecasts of *long-run* potential respond similarly to nowcasts of potential GDP addresses one possible issue raised by Blanchard (2017), namely, that demand shocks might have transitory effects on potential output. This can occur even in standard models through a number of channels, such as lower levels of physical capital following periods of disinvestment or lower levels of human capital after extended unemployment stretches. But in these models, demand shocks would still have only transitory effects on potential output, so forecasts of *long-run* potential output should remain unchanged after demand shocks, even if contemporaneous levels of potential were responding to these shocks. The fact that both nowcasts and long-run forecasts of potential respond to demand shocks suggests that the mechanism emphasized by Blanchard (2017) is not driving these results.

In short, we document a systematic response of estimates of potential GDP to shocks that have only cyclical effects on GDP. Furthermore, even some supply shocks have contradictory effects on estimates of potential GDP, in the sense that changes in the latter after oil supply shocks speak little to actual long-run changes in output. Thus, seeing ex-post that declines in GDP seem to be accounted for by changes in potential GDP, as has been the case in the U.S. since the Great Recession, says little about whether the decline in output is likely to persist or can be reversed by standard countercyclical policies.

II.D. Explaining Patterns in Impulse Responses

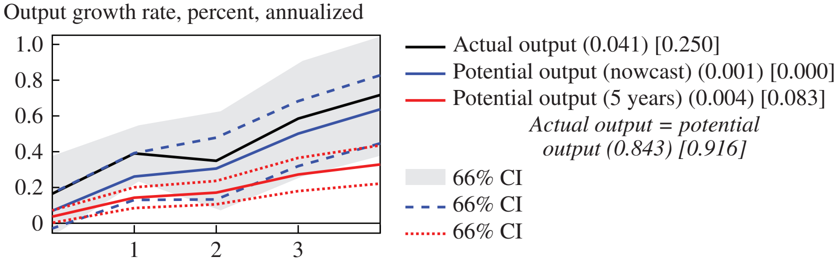
Why are estimates of potential GDP responding to shocks that only have cyclical effects, such as monetary policy and government spending shocks? One possibility is that policy institutions and statistical agencies perceive these shocks as affecting current levels of potential output (for example, if they affect current capital stocks) but not long-run levels of potential output (as would be implied by, for example, monetary neutrality). This is unlikely to be the case, however, because the long-horizon CBO forecasts of potential GDP respond about as much as their nowcasts of potential GDP.

An alternative possibility is that these estimates are relying to a large extent on simple statistical methods to measure trend (potential) levels from actual GDP. As illustrated in figure 3, one can come close to replicating the real-time Greenbook estimates of potential GDP growth by using a one-sided HP-filter on real-time GDP data available each quarter or by taking a simple, one-sided moving average of recent GDP outcomes. Because these types of methods fail to identify the different potential sources of

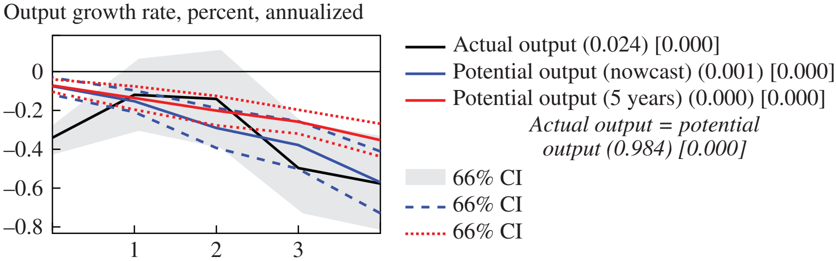
Figure 6. Responses of U.S. Output and the CBO's Estimates of Potential U.S. Output to Shocks[a]

Panel A: Supply Shocks

Total factor productivity shock (Fernald 2012)



Tax shock (Romer and Romer 2010)



Oil supply shock (Kilian 2009)

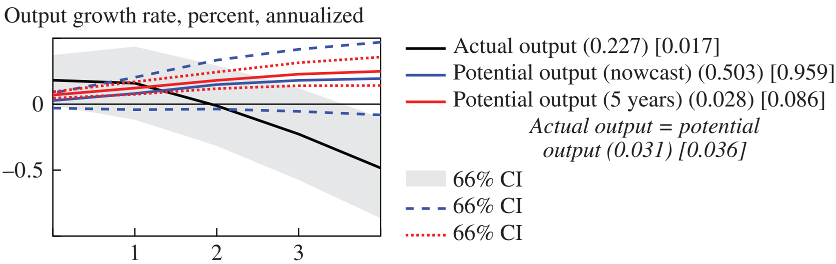
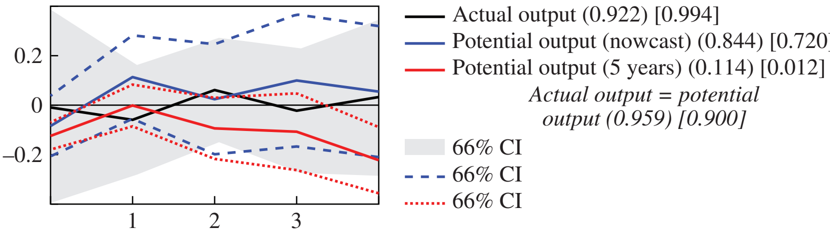


Figure 6. Responses of U.S. Output and the CBO's Estimates of Potential U.S. Output to Shocks[a] (*Continued*)

Panel B: Demand Shocks

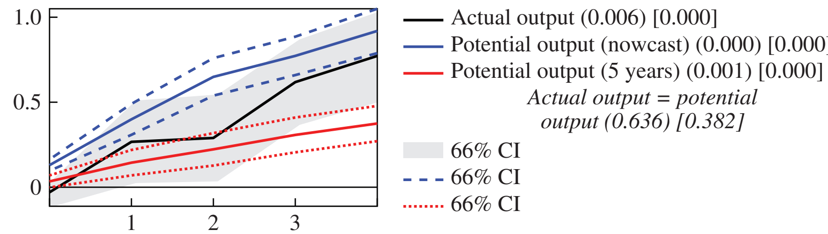
Monetary policy shock (Romer and Romer 2004)

Output growth rate, percent, annualized



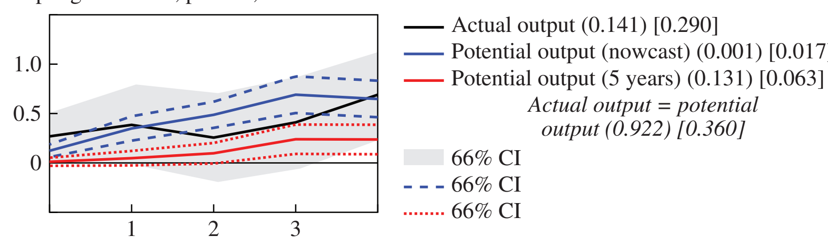
Military spending shock (Ramey 2016)

Output growth rate, percent, annualized



Government spending shock (Auerbach and Gorodnichenko 2012a, 2012b)

Output growth rate, percent, annualized



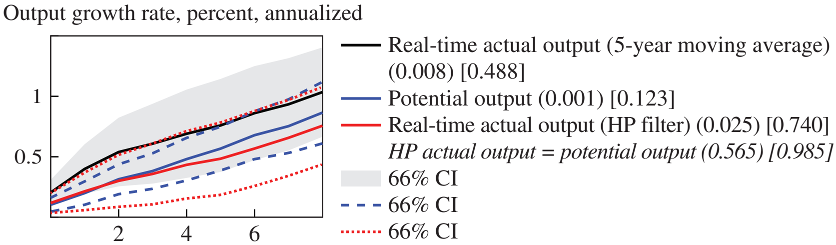
Sources: Authors' calculations, with potential output from Federal Reserve Greenbooks and identified shocks from Fernald (2012); Romer and Romer (2004, 2010); Kilian (2009); Ramey (2016); Auerbach and Gorodnichenko (2012a, 2012b).

a. This figure reports impulse response functions (IRFs) estimated using equations 2 and 3. The estimation sample covers the longest possible period with nonmissing observations for shocks and potential output (output gap) available from the Congressional Budget Office. In parentheses, we report the p values for a test of whether the response of actual (potential) output is different from zero at the maximum horizon (8 quarters). In square brackets, we show the p values for a test of whether the path of the response of actual (potential) output is different from zero over the entire duration of the IRF. The last row of the legend—for which there is no line in the graphs—reports p values for a test of equality of responses of actual and potential output (nowcast) at the maximum horizon (parentheses) and for a test of equality of the paths of the responses for actual and potential (nowcast) output are equal across horizons. CI = confidence interval.

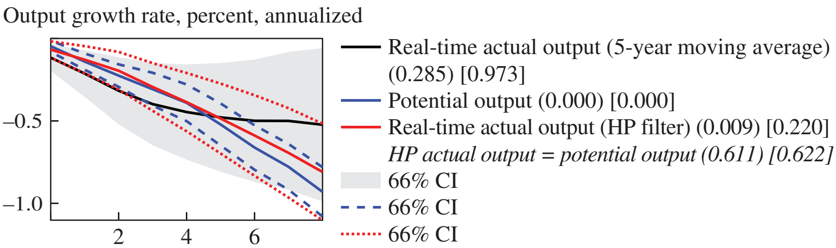
Figure 7. Responses of Greenbook Estimates of Potential U.S. Output and HP-Filtered Output to Shocks[a]

Panel A: Supply Shocks

Total factor productivity shock (Fernald)



Tax shock (Romer and Romer 2010)



Oil supply shock (Kilian 2009)

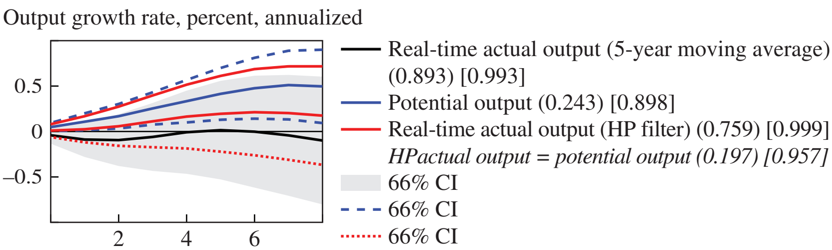
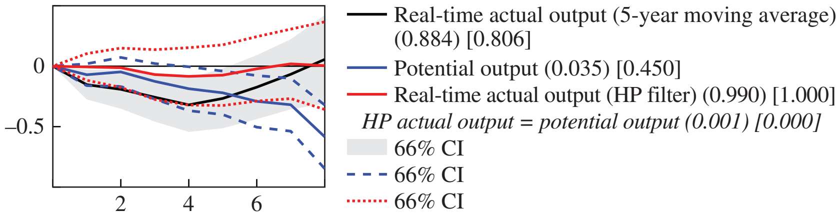


Figure 7. Responses of Greenbook Estimates of Potential U.S. Output and HP-Filtered Output to Shocks[a] (*Continued*)

Panel B: Demand Shocks

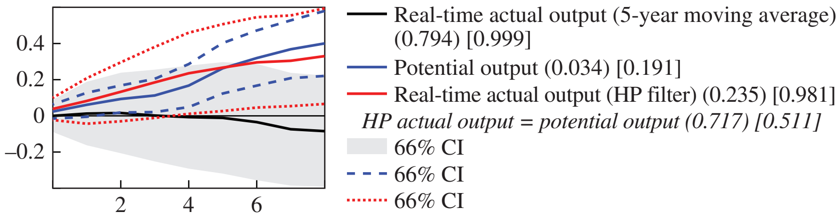
Monetary policy shock (Romer and Romer 2004)

Output growth rate, percent, annualized



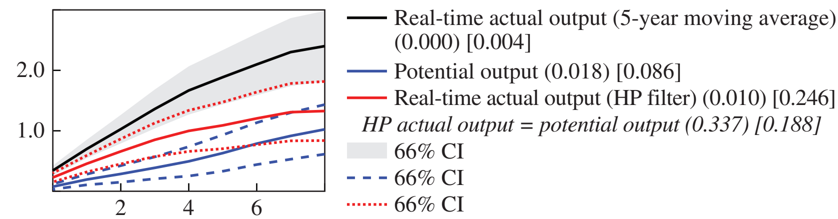
Military spending shock (Ramey 2016)

Output growth rate, percent, annualized



Government spending shock (Auerbach and Gorodnichenko 2012a, 2012b)

Output growth rate, percent, annualized



Sources: Authors' calculations, with potential output from Federal Reserve Greenbooks and identified shocks from Fernald (2012); Romer and Romer (2004, 2010); Kilian (2009); Ramey (2016); Auerbach and Gorodnichenko (2012a, 2012b).

a. This figure reports impulse response functions (IRFs) estimated using equations 2 and 3. The estimation sample covers the longest possible period with nonmissing observations for shocks and potential output (output gap) available at the Federal Reserve Bank of Philadelphia. HP-filtered (HP = Hodrick-Prescott) actual output for a given quarter is calculated as the value of the HP-filter trend for the quarter given the first vintage of GDP data that covers the given quarter. The smoothing parameter for the HP filter is set at 500,000. The five-year moving average (MA) actual output for a given quarter is calculated as the 20-quarter MA running on the current quarter and the preceding 19 quarters reported in the first vintage of GDP data that covers the given quarter. In parentheses, we report the *p* values for a test of whether the response of actual (potential) output is different from zero at the maximum horizon (eight quarters). In square brackets, we show the *p* values for a test of whether the path of the response of actual (potential) output is different from zero for all horizons of the IRF. The last row of the legend—for which there is no line in the graphs—reports *p* values for a test of equality of responses of actual and potential output at the maximum horizon (parentheses) and for a test of equality of the paths of the responses for actual and potential output are equal across horizons. CI = confidence interval.

changes in economic activity, they would naturally lead to slow-moving dynamic responses to *all* economic shocks that move actual output.

To assess this possibility, we replicate our baseline impulse responses using the same two statistical approaches to estimating potential GDP as in figure 3. In the first case, we apply a one-sided HP filter with smoothing parameter $\lambda = 500{,}000$ to real-time data on GDP. In the second, we take a five-year moving average of real GDP using real-time data. We present the results, along with the responses of potential GDP as measured by the Greenbooks in figure 7 (which is on the two preceding pages; also see the *p* values included in panel C of online appendix table 2). When using the HP-filtered series, we can very closely replicate the response of estimated potential GDP after every shock.[13] With the moving average, the fit is not as strong. The very close fit of the impulse responses using the HP filter, as well as how closely one can reproduce the unconditional time series of historical estimates of potential GDP in figure 3 with an HP-filtered series, suggests that Greenbook estimates of potential GDP incorporate little additional information relative to this purely statistical approach to estimating potential GDP.[14] It is then quite natural for these series to respond to all shocks that affect GDP, even if these movements are transitory. But this endogenous response to cyclical shocks should not be interpreted as reflecting permanent effects of these shocks on output but rather as a mechanical reaction based on how estimates of potential GDP are constructed. Equivalently, observing a downward revision in real-time estimates of potential GDP is *not* informative about whether the associated declines in actual GDP are likely to be sustained.

Another way to see how closely the HP filter can mimic real-time estimates of potential GDP, along with the potential dangers of doing so, is

13. The fact that we can match the increase in estimated potential output after an oil supply shock with the HP filter points toward a possible identification issue with these shocks. They are identified from a three-variable VAR of oil production, global economic activity (measured using an index of shipping prices), and oil prices. If oil prices are disproportionately sensitive to U.S. output (rather than global output) or shipping prices are an otherwise imperfect measure of global activity, then one might observe identified oil supply shocks disproportionately happening after sustained U.S. economic expansions (because oil prices and production are endogenous). This could lead an HP filter of real GDP to rise after an oil supply shock.

14. The best match of HP-filtered series comes with high values of λ (we use λ = 500,000). This high value is consistent with a low pass filter that allows only low frequencies with periods of about 10 years and higher. Lower values do not replicate Greenbook measures of potential GDP as closely, as can be seen in online appendix figure 6. Similarly, with moving-average measures, we can better replicate the dynamic response of Greenbook estimates of potential when averaging over long periods (10–20 years) than over shorter horizons (3–5 years), as illustrated in online appendix figure 5.

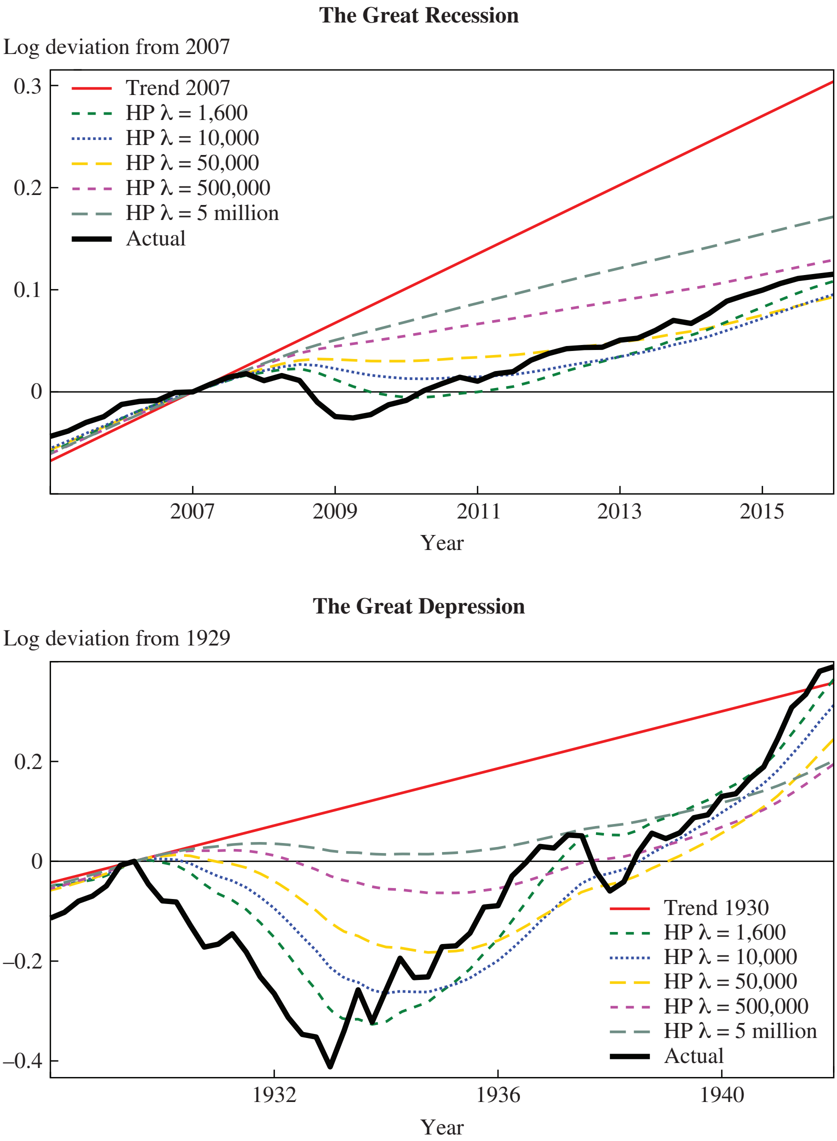
illustrated in figure 8. In the top panel, we plot the time path of potential GDP that would have been estimated in real time using the HP filter during the Great Recession. Specifically, for each quarter, we apply an HP filter to the available data and extract the trend level for that period. We then plot the sequence of these estimates over time, thereby showing the evolution of the implied real-time trend level of GDP during this historical episode for different values of the smoothing parameter. Regardless of the smoothing parameter, estimates of real-time trend output from an HP filter exhibit a significant downward revision (the magnitude of the revision declines in λ), much like the real-time estimates of official organizations in the United States, providing another illustration of how closely one can reproduce historical real-time estimates of potential output using a simple statistical filter. The danger of doing so is illustrated in the bottom panel of figure 8, which replicates this exercise for the Great Depression using data from Ramey and Sarah Zubairy (2018). The use of an HP filter to estimate potential GDP in real time over the course of the Great Depression would have implied that the output gap closed sometime between 1934 and 1936, depending on the smoothing parameter. But as illustrated in figure 8, GDP surged thereafter and real-time estimates of potential GDP began to climb back up. Unless one is prepared to entertain the idea that the Great Depression reflected negative supply shocks that were offset by positive supply shocks in the middle to late 1930s, we interpret this experience as illustrating the potential pitfalls of relying on simple statistical filters to make inferences about potential output during long-lived downturns.[15]

III. Cross-Country Evidence on the Incorporation of Shocks into Estimates of Potential

The Great Recession was of course not limited to the United States, and the persistence of output declines in most major advanced economies has also been associated with declines in their potential output, as documented by Ball (2014). Indeed, despite widespread lackluster growth by historical

15. Papell and Prodan (2012) analyze large recessions in the United States and other countries using long samples. Consistent with our analysis of the Great Depression, they find that actual output eventually catches up with prerecession projections of potential output. Gordon and Krenn (2010) document that using a bandpass filter to estimate potential GDP during the Great Depression would similarly imply implausible declines in potential between 1929 and the mid-1930s.

Figure 8. Real-Time Estimates of U.S. Potential GDP from an HP Filter during the Great Recession and Great Depression[a]



Sources: The data in the top panel are authors' calculations using underlying GDP data from FRED, the database of the Federal Reserve Bank of Saint Louis. The data in the bottom panel are from Ramey and Zubairy (2018).

a. This figure reports estimates of trend (potential output) generated by the one-sided Hodrick-Prescott (HP) filter for various values of the smoothing parameter λ. The filter is recursively applied to the final vintage of the data. For example, an estimate for 2008:Q1 uses data only up to 2008:Q1, an estimate for 2008:Q2 uses data only up to 2008:Q2, and so on.

standards since the Great Recession, the World Bank recently estimated that advanced economies have on average an output gap of zero, indicating that the large downward revisions to potential output estimated by the CBO for the U.S. since 2007 also extend to other advanced economies (World Bank 2018). To what extent can the cyclical patterns documented above in estimates of potential GDP be generalized to other countries? In this section, we turn to cross-country estimates of potential GDP, from both international organizations and professional forecasters. Using international data gives us many more observations and thus more statistical precision and power.

III.A. IMF and OECD Estimates of Potential GDP

We consider first estimates of potential GDP from two international organizations, the IMF and the OECD. Both provide estimates of the level of potential GDP for a wide range of countries.[16]

We follow the same strategy as with the U.S. and compare impulse responses of actual GDP and estimates of potential GDP from each of these two organizations to different economic shocks. However, because time samples are much shorter for most countries, we pool data across all countries in our sample. In short, for each identified shock ϵ, we estimate the following specifications:

$$(5) \qquad \Delta \log Y_{j,t} = \alpha_j + \gamma_t + \sum_{k=0}^{K} \phi_k \epsilon_{j,t-k} + error_{t,j}, \text{ and}$$

$$(6) \qquad \Delta \log Y^{*}_{j,t|t} = \delta_j + \kappa_t + \sum_{k=0}^{K} \psi_k \epsilon_{j,t-k} + error_{t,j},$$

where j indicates the country and α_j, δ_j and γ_t, κ_t respectively denote country and time fixed effects. The time frequency is semiannual, as determined by the frequency of real-time estimates of potential GDP by both the IMF and OECD.

Because of more limited data availability across countries, we cannot identify as many shocks and in the same way as was done for the United States. For productivity, we use innovations in labor productivity, after conditioning on past changes in labor productivity as well as country and

16. We exclude Norway from our analysis because this country relies heavily on energy exports.

time fixed effects.[17] For oil shocks, we continue to use Kilian's measure of oil supply shocks but interact it with a country-specific measure of oil sufficiency, from the International Energy Agency's World Energy Statistics and Balances (IEA 2017) to distinguish it from the time fixed effects.[18] For monetary policy shocks, we run a VAR for each country on GDP growth, unemployment, inflation, and the interest rate and apply a Choleski decomposition on this ordering to recover country-specific interest rate shocks. The VAR has four lags using quarterly data from 1980:Q1 until 2016:Q4 or as available.[19] Finally, fiscal shocks are differences between ex-post government spending and ex-ante forecasts of government spending from the OECD, following Auerbach and Gorodnichenko (2012a).

Turning first to the OECD sample of countries and estimates of potential GDP, figure 9 presents responses of both GDP and potential to each of the four shocks (the p values for the same tests discussed in section II are included in the figure and summarized in online appendix table 3). All four shocks yield the expected changes in GDP. Productivity shocks have an immediate and permanent effect on output, while oil supply shocks have a negative albeit delayed persistent effect on output. Both demand shocks have transitory effects on GDP which start dissipating in about one or one and a half years and are mostly gone after three years (we only show IRFs up to four semesters in the figure).

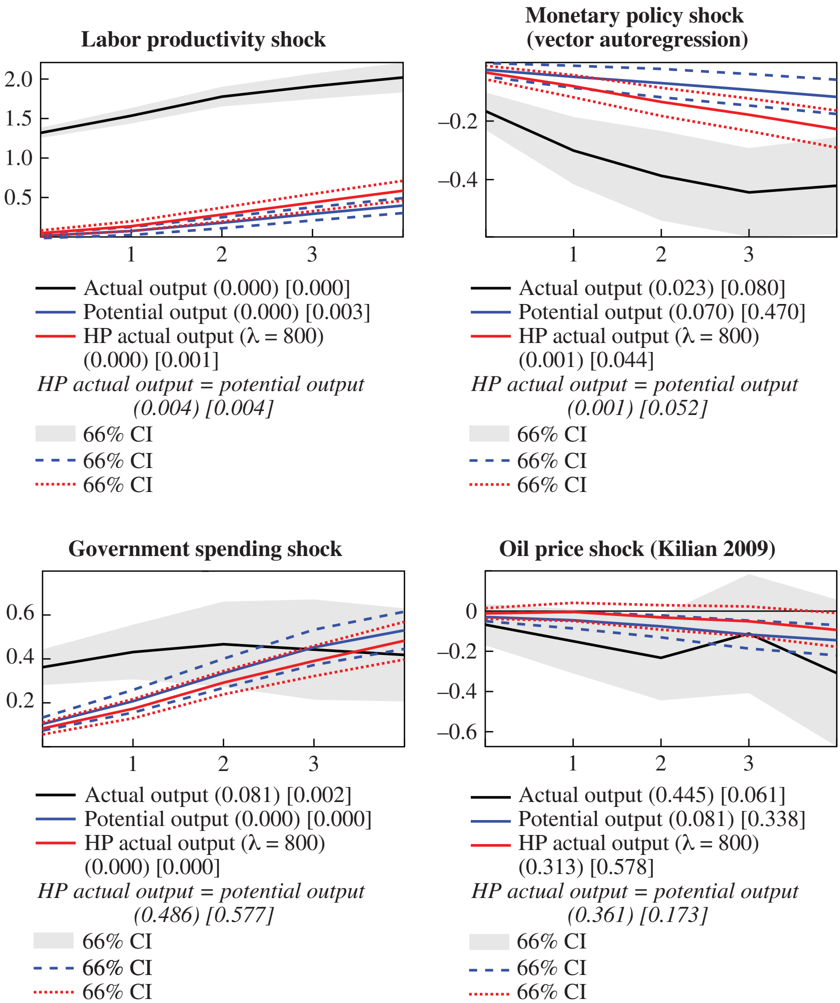
The effects of these shocks on potential GDP are consistent with those obtained for the United States. In response to productivity shocks, estimates of potential GDP evolve gradually in the direction of actual changes in output. After oil supply shocks, estimates of potential GDP decrease

17. Specifically, we use a measure of labor productivity at the semiannual frequency taken from the OECD and then regress it on lags of itself in a panel regression with country and time fixed effects, allowing coefficients on the lags of labor productivity to vary over countries, as well as a dummy for Ireland in 2015 due to its very big outliers in productivity changes. It is important to notice that this OECD measure of labor productivity is highly correlated with other measures of productivity, such as multifactor productivity from the OECD or productivity from EU-KLEMS data.

18. Oil sufficiency measures what percentage of total oil usage can be satisfied from each country's supply. Hence it ranges from 0 (if the country has no oil supply at all—for example, Belgium), passing through 1 (if the country can exactly satisfy its oil demand—for example, Australia), up to high numbers like 20 (if the country is a net exporter of oil).

19. A group of countries is in the euro zone after 1999. For these countries, we construct monetary policy shocks as follows. For the pre-euro period, we run a country-specific VAR and obtain monetary policy as described in the text. For the euro period, we run a VAR with variables measured at the level of the euro zone. From this VAR, we obtain monetary policy shocks, which we append to the shocks identified in the pre-euro period.

Figure 9. Response of the U.S. Growth Rate for Actual Output and the OECD's Measure of Potential Output (Nowcast)[a]



Sources: Authors' calculations, with potential output from the OECD and identified shocks described in the text.

a. The figure shows impulse response functions (IRFs) for growth rates of actual and potential output (nowcast). IRFs are estimated using equations 5 and 6. The horizontal axis measures time in semesters (six months). The vertical axis measures growth rate of output per year. In parentheses, we report the *p* values for a test of whether the response of actual (potential) output is different from zero at the maximum horizon (eight quarters). In square brackets, we show the *p* values for a test of whether the path of the response of actual (potential) output is different from zero across all horizons of the IRF. The last row of the legend—for which there is no line in the graphs—reports *p* values for a test of equality of responses of actual and potential output at the maximum horizon (parentheses) and for a test of equality of the paths of the responses for actual and potential output are equal across horizons. HP = Hodrick-Prescott filter. CI = confidence interval.

slightly, but this response is very weak. After both demand shocks, estimates of potential GDP gradually and persistently evolve in the same direction as the short-run changes in GDP even though these changes in GDP are transitory. Thus, we observe both the undercyclicality after productivity shocks and the overcyclicality after demand shocks documented in the United States.

Furthermore, we include in the figure the impulse response of HP-filtered real GDP (constructed for each country using real-time data and a one-sided filter) to each shock. As was the case with the United States, we find that HP-filtered GDP responds almost identically to each shock as the OECD's estimates of potential GDP. As was the case with the Greenbook estimates of potential GDP, OECD estimates do not appear to capture much more information than what is embodied in a simple univariate filter of real-time actual GDP growth rates, which can account for why their estimates of potential GDP growth rates therefore respond to shocks that have only cyclical effects on GDP.

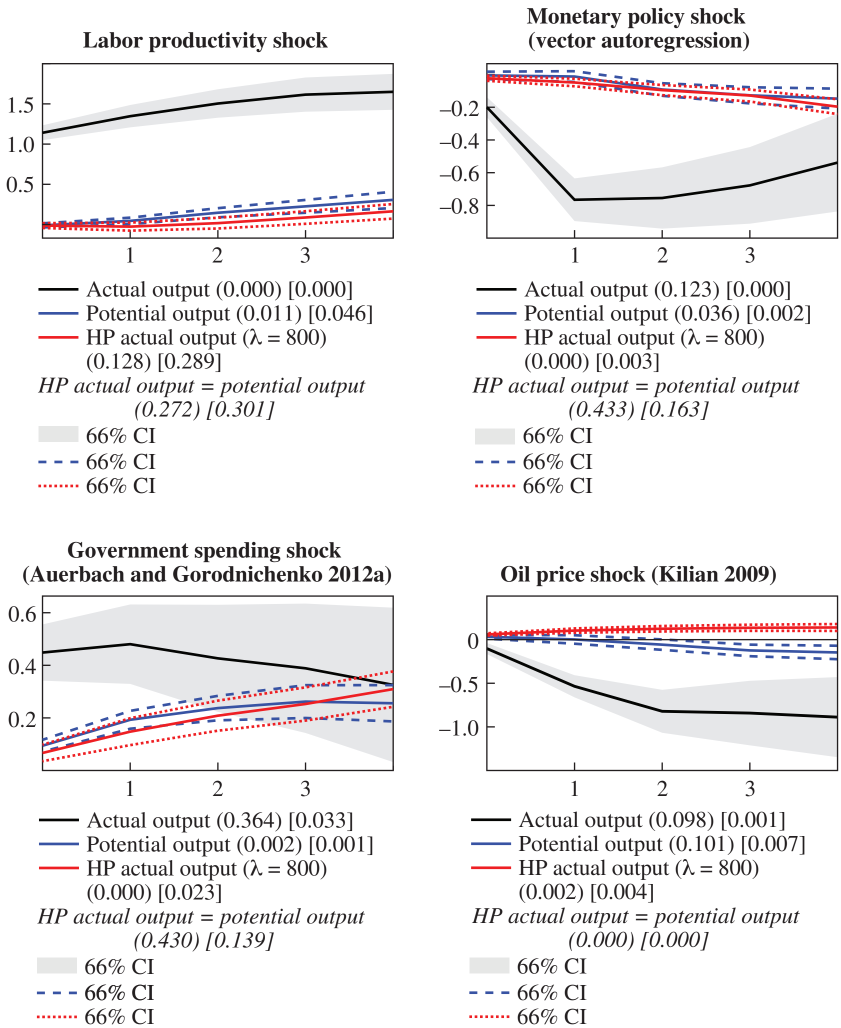
In figure 10, we produce equivalent results for the IMF sample of countries and IMF estimates of potential GDP. Despite the different countries in the sample, the estimated effects of the shocks on actual GDP are very similar to those found in the OECD sample. The responses of the IMF's estimated levels of potential GDP respond similarly to those from the OECD: They rise inertially after productivity shocks, and also respond inertially after monetary and fiscal shocks, in the same direction as the short-run response of GDP. Their response after oil supply shocks is equally weak. For comparison, we also again include responses of real-time, HP-filtered output and find, as with the OECD, that these very closely track the IMF estimates of potential output after shocks, with the only exception again being oil supply shocks.

Overall, the evidence from these two international organizations closely aligns with previous evidence from the United States: Their estimates of potential GDP are well approximated by an HP filter applied to real-time data and therefore seem to respond mechanically to short-run changes in GDP, regardless of the underlying source of economic variation. This suggests that observing revisions in one of these organizations' estimates of potential GDP in a country tells us little about how persistent the concurrent changes in GDP are likely to be.

III.B. Private Long-Horizon Forecasts of the GDP Growth Rate

In addition to forecasts from international policy organizations, we consider how private forecasters adjust their beliefs about the long-run GDP

Figure 10. Response of the U.S. Growth Rate for Actual Output and the IMF's Measure of Potential Output (Nowcast)[a]



Sources: Authors' calculations, with potential output from the IMF and identified shocks described in the text.

a. The figure shows impulse response functions (IRFs) for growth rates of actual and potential output (nowcast). IRFs are estimated using equations 5 and 6. The horizontal axis measures time in semesters (six months). The vertical axis measures growth rate of output per year. In parentheses, we report the p values for a test of whether the response of actual (potential) output is different from zero at the maximum horizon (eight quarters). In square brackets, we show the *p* values for a test of whether the path of the response of actual (potential) output is different from zero across all horizons of the IRF. The last row of the legend—for which there is no line in the graphs—reports *p* values for a test of equality of responses of actual and potential output at the maximum horizon (parentheses) and for a test of equality of the paths of the responses for actual and potential output are equal across horizons. CI = confidence interval.

growth rate in response to shocks. Although forecasts of potential GDP are not readily available, Consensus Economics provides forecasts of GDP at long horizons on a semiannual basis. To the extent that cyclical fluctuations in GDP should be complete within five years or so, these long-horizon forecasts should be equivalent to forecasts of potential GDP growth at the same horizon.

Using the same shocks as those used with the OECD and IMF samples, we replicate our previous results using private forecasts of long-run GDP for the 12 countries for which we have these forecasts (see online appendix table 1 for the countries and periods included in this sample). With the different sample of countries and time periods, the impulse responses of actual GDP are broadly similar (figure 11), although the output responses to monetary shocks are more persistent while the response to oil supply shocks is much less precise.

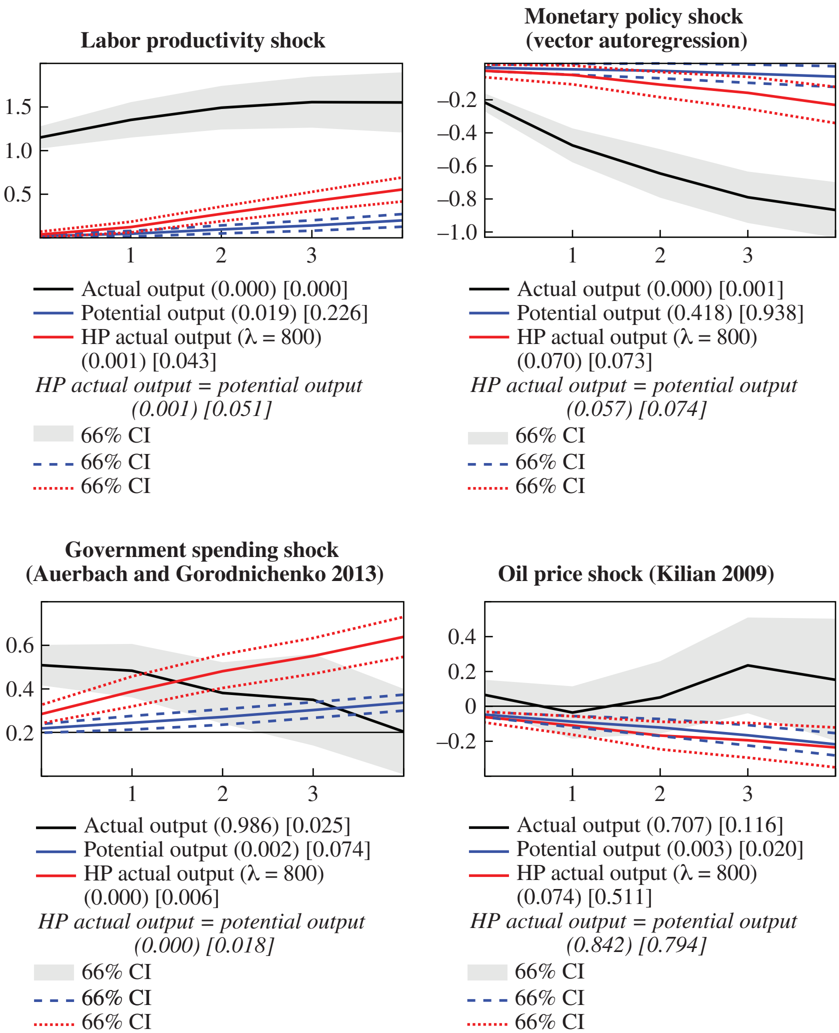
After productivity shocks, private forecasts gradually evolve in the same direction as actual output, therefore replicating the pattern observed with forecasts from public and international organizations. After the two demand shocks, the private sector forecasts also gradually evolve in the direction of the short-run movements in GDP, although the response after monetary shocks is not significant at standard levels. With respect to oil supply shocks, private forecasts of long-run GDP decline gradually.

For comparison, we also plot the implied response of HP-filtered levels of output to the same shocks and countries. For all shocks, HP-filtered forecasts evolve in the same direction as private forecasts but more rapidly. This is in contrast to what was found with estimates of potential from public and international organizations, when the estimates of potential GDP were almost identical in the impulse responses to those of an HP-filtered level of output. The more inertial response of private forecasters could reflect less rapid information updating or a difference in forecasting horizon (private forecasts are for long-run levels of GDP rather than current estimates of potential GDP).

IV. Alternative Approaches to Estimating Potential Output

The apparent inability of available estimates of potential output to differentiate between shocks that have permanent effects and those with only transitory effects raises the question of whether alternative approaches might be better. Obviously, this is a challenging task, and developing a single satisfactory method is beyond the scope of the paper. However, we can utilize available tools to get a glimpse of what may constitute

Figure 11. Response of the U.S. Growth Rate for Actual Output and Consensus Economics' 6- to 10-Year-Ahead Forecast for Actual Output[a]



Sources: Authors' calculations, with potential output from the IMF and identified shocks described in the text.

a. The figure shows impulse response functions (IRFs) for growth rates of actual and potential output (nowcast). IRFs are estimated using equations 5 and 6. The horizontal axis measures time in semesters (six months). The vertical axis measures growth rate of output per year. In parentheses, we report the *p* values for a test of whether the response of actual (potential) output is different from zero at the maximum horizon (eight quarters). In square brackets, we show the *p* values for a test of whether the path of the response of actual (potential) output is different from zero across all horizons of the IRF. The last row of the legend—for which there is no line in the graphs—reports *p* values for a test of equality of responses of actual and potential output at the maximum horizon (parentheses) and for a test of equality of the paths of the responses for actual and potential output are equal across horizons. CI = confidence interval.

a basis for a satisfactory method to estimate potential output. Specifically, we first use Blanchard and Quah's (1989) approach, designed specifically to separately identify supply and demand shocks, to show that long-run restrictions may provide a practical solution to some of the issues we have identified above. We show that this approach implies significantly different estimates of potential output during the Great Recession, and that alternative approaches yield similar conclusions.

IV.A. Blanchard and Quah Approach to Estimating Potential Output

In this simple, proof-of-concept exercise, we follow Blanchard and Quah (1989; henceforth, BQ) and estimate a bivariate VAR(8), where the variables are output growth and the unemployment rate. The identifying restriction of this model is as follows: Supply-side shocks are the structural shocks that have permanent effects on the level of output, and demand-side shocks are restricted to have zero effect on the level of output in the long run. We then interpret predicted movements in output driven by supply-side shocks as capturing potential output. The restriction that only supply-side shocks have permanent effects on output is broadly consistent with the responses of output observed in figure 4 and other results in the literature, namely, that monetary and government spending shocks do not seem to have permanent effects on output (for example, Romer and Romer 2004; Ramey 2016).

Because BQ and others emphasize the importance of structural breaks, we use a rolling window of 120 quarters.[20] When applying the BQ approach, we use real-time data to ensure that our results are not driven by information that is not available to the econometrician. In a particular quarter (say 1995:Q1), we use the vintages of real output growth and unemployment rate that were available at that point in time (obtained from the Federal Reserve Bank of Philadelphia's real-time database for macroeconomists), estimate the structural vector autoregression with long-run restriction using these series, and then perform the historical decomposition on these data to recover the component of the growth rate of actual output due to supply-side shocks for the given quarter. That is, we keep only the data point that corresponds to the last quarter in a rolling-window sample. The next quarter's (1995:Q2) historical decomposition data point is going to

20. We would like the rolling window to be big for the long-run identifying restriction to work well, but at the same time we would like it to be small to minimize exposure to structural breaks. We compromise by using a rolling window of 120 quarters, but results are similar when we use alternative rolling windows, such as 80, 100, 140, or 160 quarters.

use vintages that were not available yet in 1995:Q1, and the previous quarter's (1994:Q4) historical decomposition data point used vintages that contained less information and stopped in 1994:Q4. This approach therefore uses no more information than what was available to agents in real time, making our estimates comparable to real-time estimates of U.S. potential GDP growth.

After we recover the time series of the growth rate of output due to supply shocks (that is, our estimate of potential output), we estimate regression equations 2 and 3 on actual output and our estimate of potential output. Figure 12 shows the resulting impulse responses. We find that, in contrast to the conventional estimates of potential output, our estimate strongly reacts to supply shocks and exhibits no significant sensitivity to demand shocks. Interestingly, the reaction of our estimate for potential output to a TFP shock is stronger at short horizons than the reaction of actual output. This pattern is consistent with theoretical responses in New Keynesian models, where frictions prevent actual output from an immediate adjustment to a productivity shock so that a productivity shock creates a negative output gap in the short run. Despite its simplicity, the BQ approach can therefore make progress toward resolving puzzles in the reaction of conventional estimates of potential output to identified shocks.

It is notable that real-time estimates of potential output coming from BQ do not suffer from the same issues as those found from official estimates of potential output. One interpretation of how the latter respond to shocks is that they represent the optimal outcome in the presence of noisy information; if agents cannot differentiate between supply and demand shocks in real time, then their estimates of potential should slowly respond to each kind of shock. But the fact that the BQ methodology can, *in real time*, successfully distinguish between the two kinds of shocks suggests that this is not a binding constraint on real-time analysis but rather reflects the specific methodologies used by each organization to create measures of potential output.[21]

We can also use the BQ decomposition to revisit how potential output may have changed over the course of the Great Recession. In generating

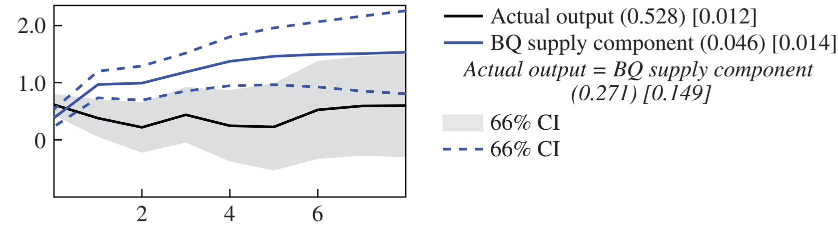
21. Another piece of evidence consistent with this interpretation is that even final (2017) estimates of potential output respond to historical supply and demand shocks in the same qualitative manner as in figure 6 (see online appendix figure 8). Despite a long delay, revised estimates of potential GDP from official agencies do not successfully distinguish between transitory and permanent shocks, suggesting that this reflects a feature of how these estimates are constructed, not an inability to distinguish between these shocks in real time.

Figure 12. Response of the U.S. Growth Rate for Actual Output and the SVAR Identified Historical Supply Component of Actual Output[a]

Panel A: Supply Shocks

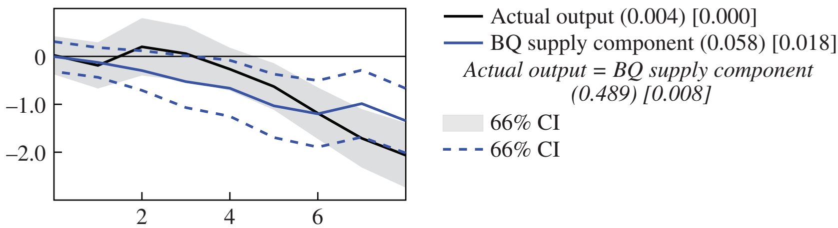
Total factor productivity shock (Fernald 2012)

Output growth rate, percent, annualized



Tax shock (Romer and Romer 2010)

Output growth rate, percent, annualized



Oil supply shock (Kilian 2009)

Output growth rate, percent, annualized

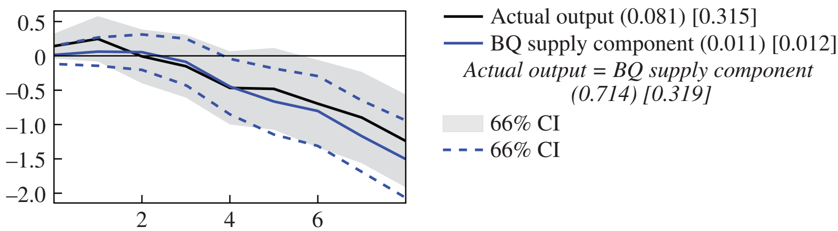
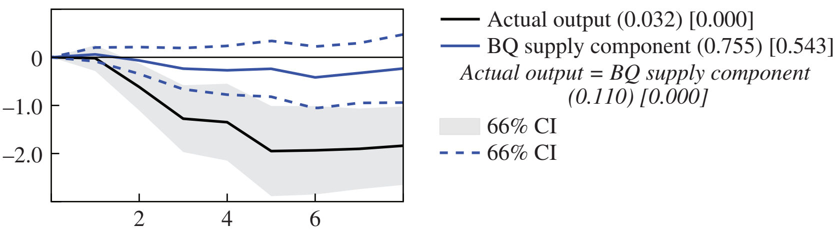


Figure 12. Response of the U.S. Growth Rate for Actual Output and the SVAR Identified Historical Supply Component of Actual Output[a] (*Continued*)

Panel B: Demand Shocks

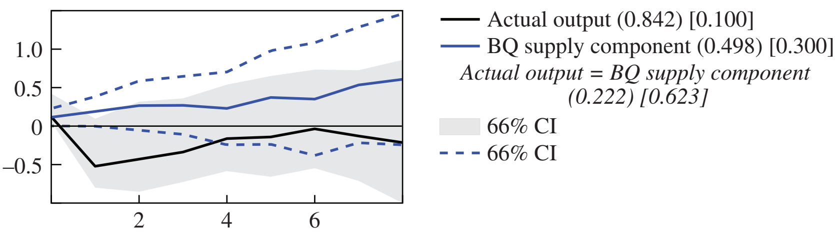
Monetary policy shock (Romer and Romer 2004)

Output growth rate, percent, annualized



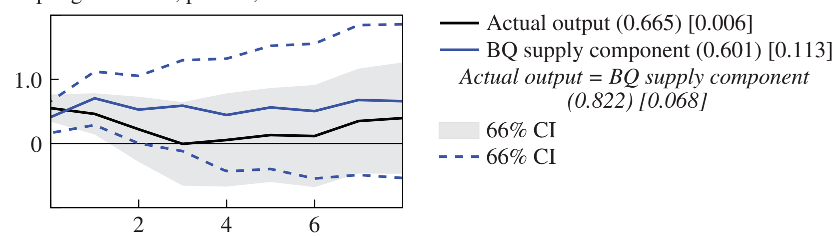
Military spending shock (Ramey 2016)

Output growth rate, percent, annualized



Government spending shock (Auerbach and Gorodnichenko 2012a, 2012b)

Output growth rate, percent, annualized



Sources: Authors' calculations, with potential output from Federal Reserve Greenbooks and identified shocks from Fernald (2012); Romer and Romer (2004, 2010); Kilian (2009); Ramey (2016); Auerbach and Gorodnichenko (2012a, 2012b).

a. SVAR = structural vector autoregression. This figure reports impulse response functions (IRFs) estimated using equations 2 and 3. The "BQ supply component" is the historical contribution of supply-side shocks—as identified by Blanchard and Quah (1989)—to the output growth rate. The estimation sample covers the longest possible period with nonmissing observations for shocks and potential output (output gap) using output gap data starting in 1970. In parentheses, we report the *p* values for a test of whether the response of actual (potential) output is different from zero at the maximum horizon (eight quarters). In square brackets, we show the *p* values for a test of whether the path of the response of actual (potential) output is different from zero across all horizons of the IRF. The last row of the legend—for which there is no line in the graphs—reports *p* values for a test of equality of responses of actual and potential output at the maximum horizon (parentheses) and for a test of equality of the paths of the responses for actual and potential output are equal across horizons. CI = confidence interval.

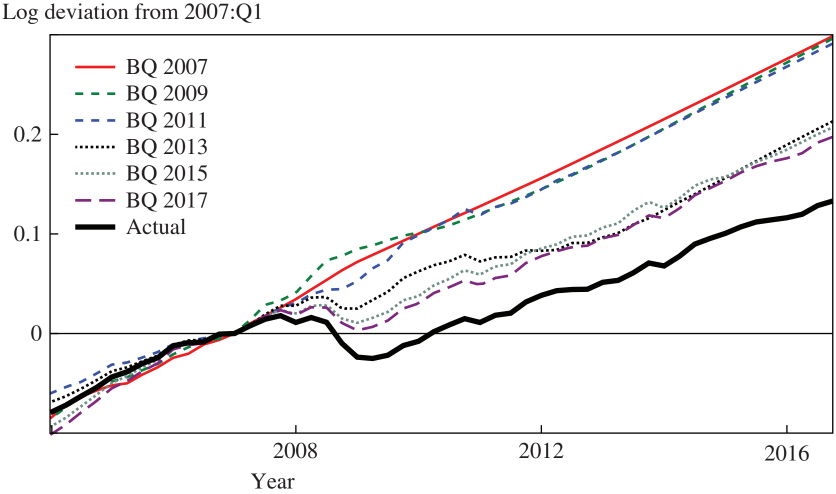
real-time estimates and forecasts of potential output using the BQ methodology, it is important to note that one must take a stand on the long-run growth rate of the economy. Heuristically, we can decompose the growth rate of output as $\Delta \log Y_t = g + \Delta \log Y_t^p + \Delta \log Y_t^c$, where g is the long-run rate of output, $\Delta \log Y_t^p$ is the growth rate of output due to "supply" shocks with permanent effects on the level of output, and $\Delta \log Y_t^c$ is the growth rate of output due to transitory "demand" shocks. We define the growth rate of potential output as $\Delta \log Y_t^* \equiv g + \Delta \log Y_t^p$. By iterating VAR coefficients from BQ forward, we construct forecasts for $\Delta \log Y_{t+h|t}^* = g + \Delta \log Y_{t+h|t}^p$, given the history of supply shocks up to period t. Then we cumulate $\Delta \log Y_{t+h|t}^*$ over $0, \ldots, H$ to compute the response of the level of potential output to a shock. Note that in this calculation, we follow BQ and assume that shocks do not influence g, the growth rate of output in the long run. Although this assumption is consistent with the fact that the growth rate of output per capita in the United States has been remarkably stable, at 2 percent a year over the last 150 years (Jones 2016), it is nonetheless an important assumption. In the context of using BQ for the Great Recession, we apply the long-run growth rate of GDP from the 1977–2007 period (3.1 percent) and assume that it remains invariant to the Great Recession.

The resulting real-time revisions in potential output using the BQ methodology during the Great Recession are plotted in the top panel of figure 13. Like official estimates, we find that there are declines in potential output during the Great Recession that take some time to uncover; the first significant downward revisions for 2009 potential output occur using the 2013 estimates. But there is little predictability in subsequent revisions; they all closely track the 2013 estimates of the path of output. And unlike the official estimates, the BQ approach points to a large and continuing gap between actual output and potential. By 2016, we estimate U.S. potential output to have grown by about 5 log percentage points more than actual output since 2007, a difference that could potentially be closed through the use of demand-side policies.

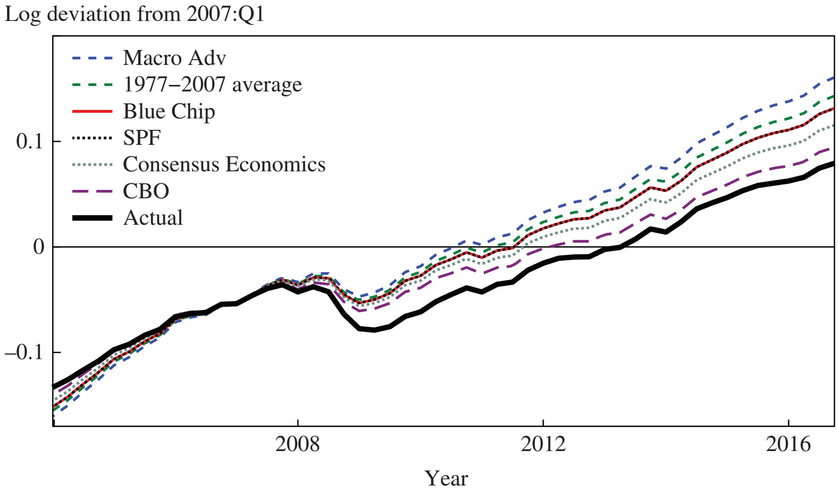
Furthermore, it is likely that BQ estimates represent an overestimate of the decline in potential output. This is because, since the onset of the zero bound on interest rates, even transitory demand shocks should be expected to have more persistent effects than they normally would, given the absence of offsetting monetary policy actions. Because the BQ approach is estimated over a long period, more persistent demand shocks during the zero lower bound are likely to be in part attributed to "supply shocks" in the BQ decomposition. Some of the estimated decline in potential output since the

Figure 13. Revisions in Potential GDP during the Great Recession from the Blanchard-Quah Methodology[a]

Assuming Long-Run Growth Rate Equal to Average 1977–2007 Value (3.1 Percent)



Using Alternative Long-Run Growth Rates, BQ 2017 Vintage



Sources: Authors' calculations, following the structural vector autoregression methodology of Blanchard and Quah (1989) and various measures of long-run growth between 2007 and 2017 from Consensus Economics, Macroeconomic Advisers, Survey of Professional Forecasters, the Congressional Budget Office, and Blue Chip Economic Forecasts.

a. The top panel plots the real-time estimates and forecasts of potential GDP, following Blanchard and Quah (1989), for different rolling windows. YYYY in "BQ YYYY" shows the last year of the rolling window. See section IV.A for details. The bottom panel plots BQ 2017 for different values of g, which are taken from the sources indicated in the legend: Macro Adv = Macroeconomic Advisers; Blue Chip = Blue Chip Economic Forecasts; SPF = Survey of Professional Forecasters; CBO = Congressional Budget Office. See section IV.B for details.

Great Recession attributed to supply-side factors is therefore likely to be transitory, making the output gap even larger than our estimates suggest.

Because of the possible sensitivity of BQ estimates of potential GDP to assumptions about the long-run growth rate, we consider a number of other values for the long-run growth rate of output that were suggested before the Great Recession. We view it as important to restrict our attention to pre–Great Recession estimates because these already include predictable deterministic changes in growth after 2007 (such as from the retirement of the Baby Boomers) but are not contaminated by the persistent changes in output since the Great Recession. Indeed, as we documented using long-run projections of professional/official forecasters in subsection III.B, real-time estimates of long-run growth respond to shocks that have only transitory effects, so we should expect these estimates to have been significantly reduced since the Great Recession (as most in fact have been), but this is not informative about whether these changes should be expected to persist.[22]

Given the difficulty inherent in making forecasts about future productivity growth, the main driver of long-run GDP growth, there was significant uncertainty about the long-run future growth rates of the United States' GDP before the Great Recession. For example, Macroeconomic Advisers, a prominent economic forecasting firm, was predicting a relatively high long-run growth rate of 3.3 percent. Many other professional forecasters were similarly optimistic, with forecasters in both the Blue Chip Economic Forecasts and the Survey of Professional Forecasters predicting long-run growth rates of 3.0 percent, just under the postwar average of 3.1 percent. Other forecasters were somewhat more pessimistic. For example, forecasters at Consensus Economics were predicting an average long-run growth rate of 2.8 percent (there was much disagreement across forecasters; the standard deviation is 0.6 percent). The CBO was even more pessimistic, predicting an average growth rate of just 2.6 percent in the long run. We show the implications of each of these assumptions for BQ decompositions since the Great Recession in the bottom panel of figure 13. Depending on the source of long-term projections, the output gap has fallen anywhere between 15 percent (Macroeconomic Advisers) to 2 percent (CBO) since the Great Recession.

22. We find similar results when we adjust output by the size of the civilian population (online appendix figure 9).

IV.B. Alternative Estimates of Potential Output since the Great Recession

Although these different estimates from the BQ methodology all imply significant remaining slackness, they also point to the difficulty of pinning down the output gap using a single procedure. In this subsection, we consider several alternative theory-based approaches to investigate the robustness of this finding.

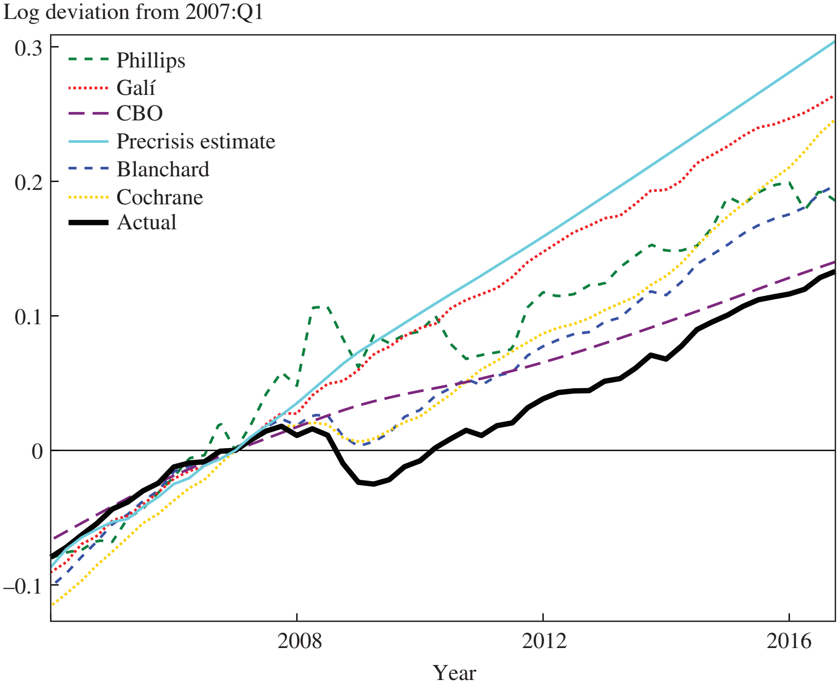
One approach closely related to BQ is from Galí (1999). He proposes identifying technology shocks in a VAR through long-run restrictions by assuming that these shocks change labor productivity in the long run while other shocks do not. We apply the same two-variable VAR as used by Galí (1999) to real-time data and define the real-time level of potential output as that coming only from the identified technology shocks. As illustrated in figure 14, this approach points to even smaller changes in potential output over the course of the Great Recession, perhaps due to the narrower interpretation of the types of shocks that affect potential output than in BQ. The 2017 level of potential output is only 5 log percentage points lower when estimated using 2017 data than forecasted from 2006 data, yielding a growth in the output gap by 2017 of well over 10 log percentage points relative to 2007.[23]

Cochrane (1994) proposes an alternative approach to identifying permanent changes in GDP by exploiting the consumption/output ratio. Under Milton Friedman's (1957) Permanent Income Hypothesis, consumption changes reflect permanent changes in income, so adding information about consumption can help decompose transitory from permanent changes in income. Applying his methodology to real-time data on consumption and GDP, and identifying potential GDP as those changes associated with changes in consumption, yields a surprisingly similar path of revisions in potential output over the Great Recession as the BQ approach, as illustrated in figure 14. As with Galí's (1999) approach, the implied output gap in 2017 is therefore more than 10 log percentage points bigger than in 2007 when applying the same long-run growth rate as in BQ estimates (3.1 percent).[24]

23. One could also follow King and others (1991), Gonzalo and Ng (2001), and others to consider VARs that include more than two variables or use other permanent/transitory decompositions.

24. We report results for different vintages of the Galí and Cochrane approaches in online appendix figure 10.

Figure 14. Alternative Approaches to Estimating Potential GDP in Real Time during the Great Recession[a]



Sources: Authors' calculations of potential output, following various methodologies (Blanchard and Quah 1989; Galí 1999; Cochrane 1994), or using information in the Phillips curve or from the Congressional Budget Office.

a. The figure plots the 2017 estimates of the path of potential GDP from these approaches, as well as the Blanchard and Quah (1989, "Blanchard") approach, the Phillips curve, the CBO's estimates for 2017, and the 2007 precrisis estimate. "Actual" denotes the path of real GDP. See section IV.B for the details.

Importantly, the Cochrane approach is immune to concerns about hysteresis, because it does not try to distinguish between supply and demand shocks based on their long-run effects. If hysteresis is present, then even transitory shocks should have effects on consumption due to their long-lived effects on income. As a result, they would be incorporated into the resulting estimates of potential output. Furthermore, this approach is also likely to overstate the decline in potential output over this time period. If some households are credit-constrained ("hand-to-mouth") and adjust their consumption to transitory income changes, then we will measure declines in potential GDP even from some transitory shocks, thereby overstating the change in potential GDP since the Great Recession and understating the current amount of economic slackness.

Closer in spirit to Okun's (1962) approach is to infer information about potential output from the inflation rate. In New Keynesian models, nominal rigidities generate an expectations-augmented Phillips curve that relates inflation to expected inflation and the output gap (or the deviation of unemployment from the natural rate of unemployment). Conditional on observing inflation, expected inflation, and real GDP, one can then use the Phillips curve to infer the potential level of GDP (under the assumption of no markup shocks). Following Coibion and Gorodnichenko (2015b), we estimate an expectations-augmented Phillips curve during the pre–Great Recession period using inflation expectations from the Michigan Survey of Consumers. As shown by Coibion and Gorodnichenko (2015b), conditioning on household forecasts of inflation yields a stable Phillips curve since the 1960s and eliminates the puzzle of the "missing disinflation" during the early years of the Great Recession. We then apply this Phillips curve to the period since the Great Recession to infer what path of potential output is required to account for inflation dynamics during this period.

A key advantage of this approach is that it does not rely on long-run restrictions, which may be sensitive to structural breaks (Fernald 2007). We plot a smoothed version of 2017 estimates of potential GDP over the period of the Great Recession in figure 14, along with the 2017 estimates from other approaches for comparison.[25] The implied potential GDP from the Phillips curve does not decline much until 2011, significantly later than other approaches. However, by 2017, the resulting estimate of potential GDP is close to that of the BQ approach, pointing to an output gap of about 5 log percentage points. In a related paper (Coibion, Gorodnichenko, and Ulate 2019), we do more extensive work using the expectations-augmented Phillips curve to back out potential output. We show that this approach works systematically across countries and that the measures of potential output that it delivers paint a similar picture to the ones obtained in this paper using the BQ approach.

In short, bringing additional information to bear on the identification of potential output—whether from labor productivity, consumption, or inflation—combined with theoretical predictions regarding how these variables relate to potential GDP, largely confirms the findings of the BQ

25. We plot a smoothed version because sampling uncertainty in inflation expectations measured by the Michigan Survey of Consumers (500 households participate in the survey in a typical month) generates high-frequency noise in estimates of potential GDP.

approach. Each approach points to nontrivial revisions in potential output since the Great Recession, but not nearly as large as those coming from the official organizations. This implies that current U.S. output likely remains significantly below potential output, and therefore that further stabilization policies could be warranted.

IV.C. Can the Output Gap Be Large When Unemployment Is Low?

Our view that a significant output gap likely remains in the United States a decade since the start of the Great Recession may seem at odds with the conclusion one might reach from looking at recent U.S. unemployment rates. For example, an output gap of 5 percent would, using Okun's law, require a negative unemployment rate gap of about 1.5 percent.[26] With the U.S. unemployment rate having fallen below 4 percent in April 2018, this would imply a natural rate of unemployment of about 2.5 percent. In contrast, typical estimates of the NAIRU point toward much higher values (the 2018 CBO estimate is 4.6 percent). Is it possible to reconcile recent labor market dynamics with our estimates of potential output? In this subsection, we argue that the answer is unambiguously yes, and that it is the alternative view—namely, that labor markets are currently very tight—that seems at odds with other economic dynamics.

First, the evidence from a number of other macroeconomic variables is consistent with the view that much economic slackness remains. Consumption dynamics, for example, suggest that permanent declines in income have been quite limited since the recession, as shown in subsection IV.B, which also documents that the behavior of inflation relative to inflation expectations is consistent with significant economic slackness remaining. Other variables point toward a very similar conclusion. For example, capacity utilization is a commonly used measure of the state of the business cycle. By the end of 2017, utilization was at 77 percent, well below its average value of 81 percent over the 1977–2007 period, with only 14 percent of quarters during that period having utilization rates of less than 77 percent. By historical standards, such low utilization rates are hard to reconcile with output being at or above its normal productive capacity. Wages also paint a picture of a labor market that remains slack: Annual nominal and real wage growth in the last quarter of 2017 were at the 21st and 6th percentiles, respectively, of the distribution of their historical

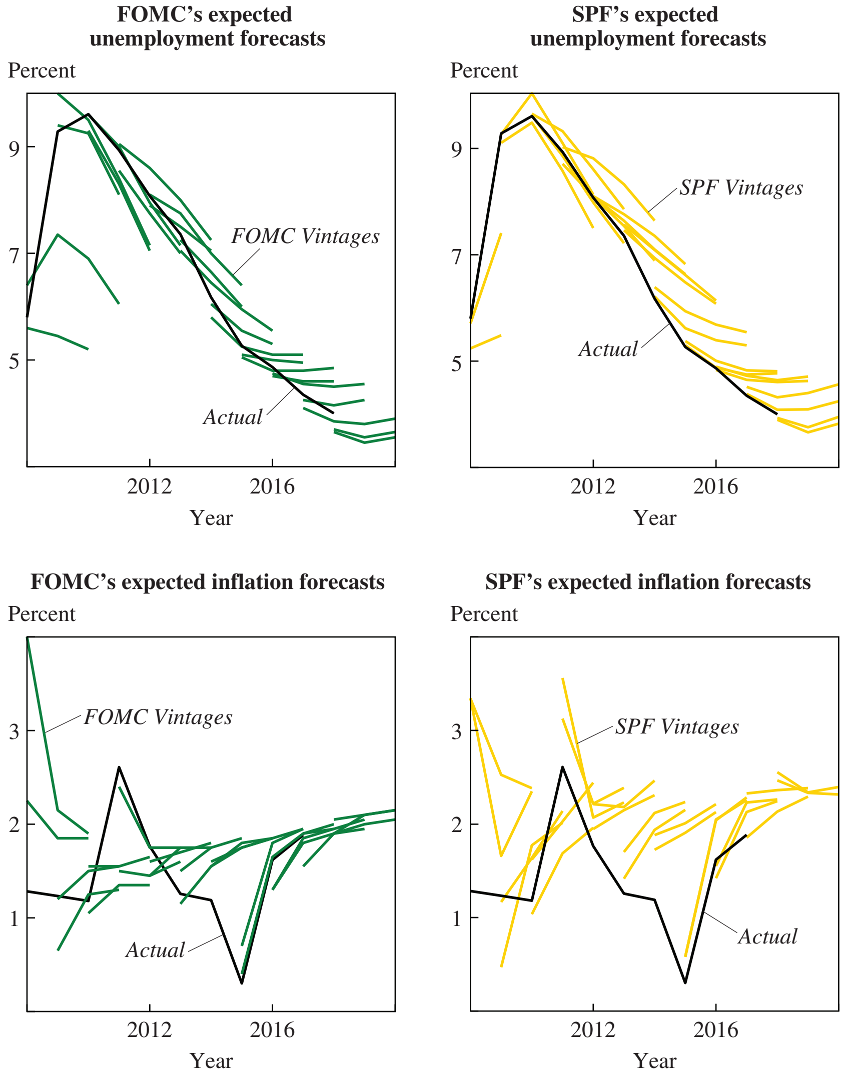
26. For all Okun's law calculations, we use a coefficient of 3, such that each change in the unemployment gap of 1 percentage point is associated with a change in the output gap of 3 percentage points (for a range of estimates of Okun's law, see Knotek 2007).

values from 1977 to 2007. Also by historical standards, it is difficult to reconcile tight labor markets with such low growth rates in wages.

Second, any statement about the natural rate of unemployment must be tentative at best, given the conceptual and measurement issues involved. Indeed, many of the same challenges as those associated with estimating the potential level of GDP are also present in estimating the natural rate of unemployment, so there is little reason to expect one to be more accurately measured than the other. Consistent with this, we observe similar patterns of systematic revisions in estimates of the natural rate of unemployment as we do in estimates of potential GDP. For example, these revisions tend to be in the direction of actual changes in unemployment, much as we observed with potential GDP. The top right panel of figure 15 plots projected unemployment rates of professional forecasters at different moments during the recovery, and their estimates of the natural rate of unemployment over time are given in figure 16. When unemployment first began to decline after its peak during the Great Recession, professional forecasters expected a gradual decline in unemployment toward a natural rate that was estimated to be nearly 6 percent. But as unemployment rates fell over time, professionals also continuously revised their estimates of the natural rate downward, with their current estimates being just above 4 percent. Importantly, professional forecasters have been consistently too pessimistic in their unemployment projections since 2011. The CBO's estimates of the natural rate of unemployment have followed an identical pattern, albeit with smaller changes (figure 16). The top left panel of figure 15 shows that FOMC members have similarly adjusted downward the levels toward which they project unemployment rates will converge, though they do not publicly provide explicit forecasts of the natural rate of unemployment.

Third, predictions about nominal variables based on perceptions of a tightening labor market have been significantly off target in recent years. As described in subsection IV.B, an expectations-augmented Phillips curve requires a significant output gap to account for inflation dynamics since the Great Recession. But even without imposing an expectations-augmented Phillips curve, forecasts based on tight labor markets have failed to adequately predict inflation. For example, the bottom right panel of figure 15 plots inflation forecasts from the Survey of Professional Forecasters over the course of the Great Recession; these have repeatedly overpredicted inflation since 2013, consistent with professionals overestimating the tightness in labor markets. A similar pattern is visible using inflation forecasts from the FOMC members over the same period (the bottom left panel of figure 15). The degree of overestimation of inflation is more limited in

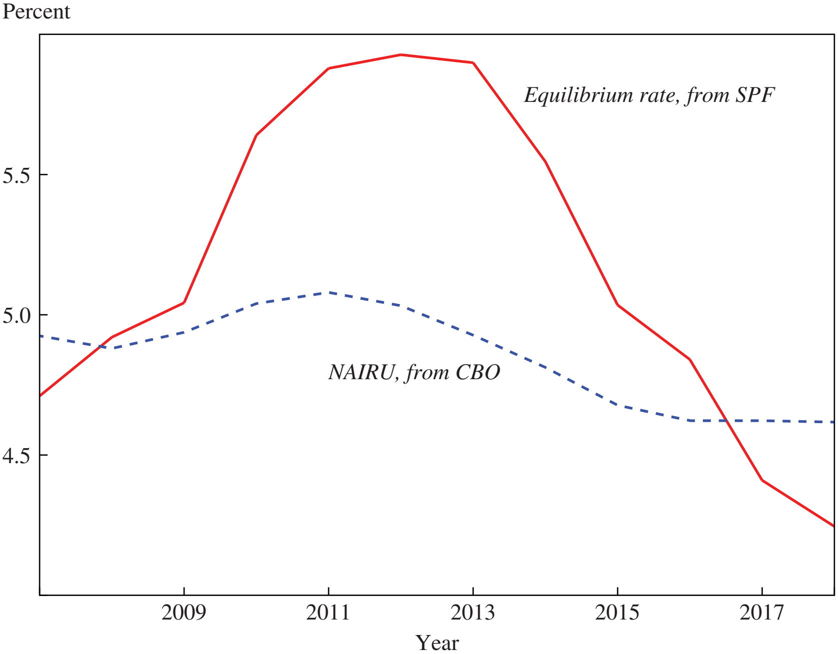
Figure 15. Unemployment and Inflation Forecasts since the Great Recession[a]



Sources: FRED, the database of the Federal Reserve Bank of Saint Louis; Federal Open Market Committee; Survey of Professional Forecasters.

a. This figure plots actual unemployment rates (top panels) and inflation rates (bottom panels), as well as projected rates reported by the Survey of Professional Forecasters (SPF) and from surveys of members of the Federal Open Market Committee (FOMC).

Figure 16. Estimates of the Natural Rate of Unemployment[a]

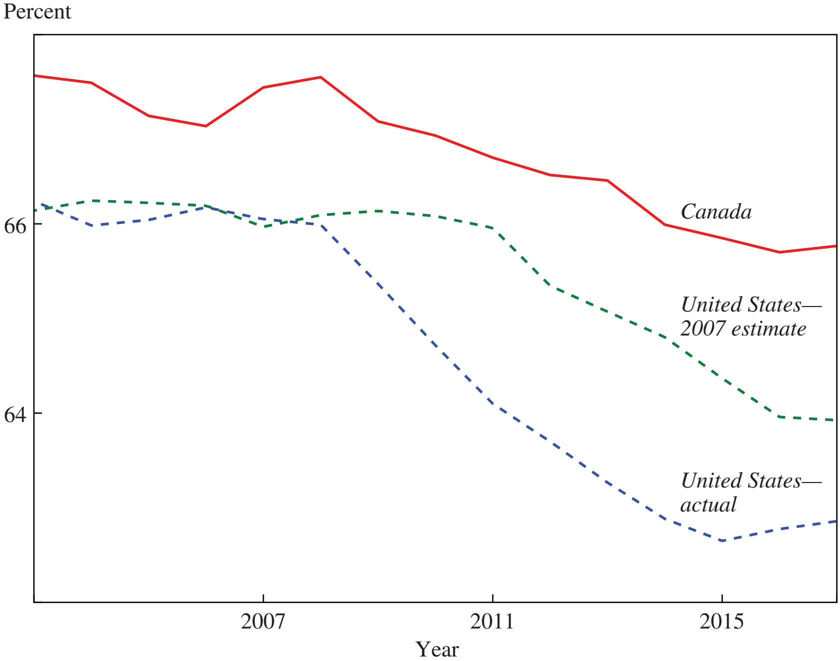


Sources: Congressional Budget Office; Survey of Professional Forecasters.
a. This figure shows time series of the nonaccelerating inflation rate of unemployment (NAIRU), estimated by the Congressional Budget Office (CBO), and the consensus real-time estimate of the equilibrium rate of unemployment from the Survey of Professional Forecasters (SPF).

FOMC forecasts, but this likely reflects the institutional nature of these forecasts: Policymakers need to present forecasts of inflation that converge to the 2 percent target or risk casting doubt on their credibility (Tarullo 2017).

The issues with measuring tightness in labor markets extend beyond the difficulties associated with estimating the natural rate of unemployment and extend to the challenge of using the unemployment rate as a measure of slackness. In an environment where labor force participation exhibits clear business cycle variation, the unemployment rate may not be a sufficient metric of business cycle conditions. And this issue is not new; over the course of the late 1990s, for example, Federal Reserve chairman Alan Greenspan allowed unemployment to fall significantly below the then-estimated natural levels of unemployment (the “Greenspan gamble”). Instead of generating a rise in inflation, the result was an increase in labor force participation (from 66.5 percent in January 1996 to 67.3 percent

Figure 17. Labor Force Participation in Canada and the United States[a]



Sources: Statistics Canada; Congressional Budget Office; U.S. Bureau of Labor Statistics.

a. The figure plots time series of actual and projected labor force participation rates. The Canadian series is from Statistics Canada. The U.S. actual series is from the Bureau of Labor Statistics. The 10-year-ahead projection—as of 2007—for the U.S. is from the Congressional Budget Office.

in April 2000), which led the CBO to later revise downward its estimate of the 1999 natural rate of unemployment from 5.6 to 4.8 percent. This endogeneity of the labor force participation rate appears to have become increasingly pronounced since the Great Recession. It is well known that labor force participation in the U.S. has declined significantly since the start of the Great Recession relative to 2007 projections (figure 17). How much of this decline is likely to reflect an endogenous decision by some to abandon the labor force because of limited job prospects? One way to gauge this is to compare labor force participation in the United States with that in Canada, which has a similar demographic structure and trends and thus is a frequent benchmark for comparison (see, for example, Card and Freeman 1993). But Canada is also a country that did not experience a serious financial crisis or a recession anywhere near the size of what was experienced in the U.S. As illustrated in figure 17, labor force participation in Canada also declined since 2007, but by far less than in the U.S.—1.7 versus 3.2 percent. In fact, the decline in labor force participation in Canada since 2007—2.0 percent—corresponds almost exactly to the decline in participation that the CBO predicted would happen in the U.S. in 2007, before the start of the Great Recession. If we measured the 2017 U.S. unemployment rate relative to a labor force size consistent with a declining participation rate of 2.0 percent instead of 3.2 percent, we would have an estimated unemployment rate in 2017 of 5.3 percent (instead of 4.4 percent) and an output gap of 5 percent, which would imply, via Okun's law, a natural rate of unemployment of 3.7 percent.

Christopher Erceg and Andrew Levin (2014) provide another way to gauge the cyclical sensitivity of labor force participation during the Great Recession by exploiting the cross-state variation in employment outcomes. They find that states experiencing larger increases in unemployment during the Great Recession also experienced larger declines in participation in subsequent years, a feature we verify over a longer time span in online appendix figure 11. They find that each 1 percentage point of higher unemployment is associated with a 0.3 percent decline in the participation rate. Extrapolating this to the aggregate economy, the increase in the national unemployment rate by 5 percentage points between 2007 and 2009 should therefore be expected to generate a decline in participation of about 1.5 percentage points. Hence, endogenous participation can account for all the unexpected decline in the participation rate observed since the Great Recession.[27] Accounting for this change in the participation of the unemployed yields an adjusted unemployment rate of 5.8 percent for 2017 and, via Okun's law and an estimated output gap of 5 percent, a natural rate of unemployment of 4.1 percent.

This sensitivity of both the measured unemployment rate and the estimated natural rate of unemployment should give one pause when thinking about the cyclical state of the economy based on the labor market. The endogeneity of labor force participation puts typical values of both in question. Because estimates of potential output are not being normalized by

27. Erceg and Levin (2014) focus on the labor force participation rate for prime age adults. In online appendix figure 11, we present equivalent results using changes in total labor force participation from 2007 to 2017 across states. We find that an increase of 1 percentage point in the unemployment rate between 2007 and 2009 is associated with a decline of 0.15 percentage point in the labor force participation rate through 2017, or half the sensitivity found by Erceg and Levin (2014). Hence, our estimates imply that the aggregate rise in unemployment from 2007 to 2009 can account for three-fourths of the unpredictable component of the decline in participation.

an endogenous variable the way unemployment rates are, this provides another reason to focus on measuring output gaps rather than unemployment gaps. However, estimating potential output is no panacea for the measurement problems associated with labor market variables. As Okun (1962, 1) observed, "The quantification of potential output is at best an uncertain estimate and not a firm, precise measure." Indeed, estimating potential output is hard because statistical issues are magnified by sensitivity to economic assumptions. For instance, forecasts of actual output are routinely associated with wide confidence bands (for example, standard errors for the Fed and private one-year-ahead forecasts are often greater than 1 percentage point). Because potential output is aimed to project long-run dynamics, sampling uncertainty is amplified in these projections. This uncertainty is further exacerbated by using long-run restrictions, as in the BQ approach and similar methods, in relatively short samples. Structural breaks and low-frequency variation in the data add another layer of complexity.

The sensitivity of potential output estimates to variation in economic assumptions is equally humbling. For example, BQ and similar approaches assume that g, the long-run growth rate of potential output, does not respond to economic shocks; but conceivably, g may persistently react to these shocks. Because even small differences in growth rates are compounded into large magnitudes over time, a weak sensitivity of g to shocks can translate into significant variation in potential output estimates. Concretely, if we overstate g by 0.1 percent a year, over 10 years we can overstate the output gap by 1 percentage point.[28] As a result, because estimating potential output is inherently so challenging, one should interpret our estimates in this section, and indeed all estimates of the potential level of output, as tentative. This uncertainty surrounding estimates of potential output and the natural rate of unemployment implies that risk management should be a primary consideration in policymakers' decisionmaking processes.

V. Conclusion

Our results speak to two distinct but related questions. The first is how real-time estimates of potential output respond to transitory versus permanent economic shocks and therefore how we should interpret revisions in

28. The degree of uncertainty about what value to use for g is large. Gordon (2014), for example, argues that g is likely to be only 1.6 percent a year between 2014 and 2020, well under the CBO's forecast of 2.2 percent a year, and far below the historical average of 3.1 percent (1947–2017 sample).

estimates of potential output observed in the data. The second is how high-quality, real-time estimates of potential *should* react to economic shocks.

With respect to the first question, we provide robust evidence that real-time estimates of potential output respond to *all* identified economic shocks, whether transitory or permanent. Observing a sequence of revisions in estimates of potential output, like those since the start of the Great Recession, therefore tells us little about whether declines in GDP are likely to be permanent or transitory. Instead, approaches like those of Blanchard and Quah (1989), who explicitly distinguish between temporary and long-lived shocks, are much more successful in this respect. Importantly, they suggest that current U.S. GDP is significantly below its longer-run potential and therefore that the U.S. economy remains in need of ample stimulus from monetary and fiscal authorities.

In terms of how high-quality estimates of potential *should* respond to shocks, the answer is sensitive to the concept of potential output one has in mind and the purpose that it is supposed to serve. For an agency like the International Monetary Fund that is concerned with constructing cyclically adjusted balances and long-run fiscal trends, the relevant measure of potential output is precisely one that strips out cyclical variation in GDP and identifies long-run changes. Our results suggest that the current methods used by this and similar agencies are largely unsuccessful in this respect; their revisions are contaminated by transitory shocks and respond too slowly to long-lived shocks. For example, tax cuts that have immediate and permanent effects on output are not fully reflected in official estimates of potential output for several years, suggesting that the effects of tax changes on projected revenues are likely overstated. In this sense, our results are related to the research of Blanchard and Daniel Leigh (2012), who argue that the IMF underestimates the fiscal multipliers of austerity measures.

At the same time, it is important to bear in mind the severe constraints that hamper the ability of both public and private organizations to estimate potential GDP in real time. Not only are there profound statistical and economic challenges involved, as described in subsection IV.C, but tight budgetary restrictions also make the systematic creation and updating of these estimates in real time a significant challenge for public institutions. The political implications of the estimates of potential GDP created by these agencies also present additional constraints on officials' ability to experiment with alternative procedures. The objective of our paper should therefore not be interpreted as criticizing these particular organizations but rather as highlighting the limitations of the methods that are currently being relied upon for both fiscal and monetary policymaking, as well as proposing potential alternatives.

The approaches that we consider here, either because they explicitly distinguish between transitory and permanent shocks like Blanchard and Quah (1989) or incorporate additional information like consumption or inflation, can help address some of the limitations of currently used methods and lead to improved estimates of cyclically adjusted levels of GDP. It is likely that there remains much room for further improvement in the real-time measurement of potential output. One strategy would be to combine some of the different approaches used in this paper (as well as others), in the hope that combining different sources of information could augment the precision of the resulting estimates. A complementary approach might be to consider the dynamics of potential GDP jointly with the natural rate of unemployment and the natural rate of interest, concepts that are closely related but typically are estimated separately. Because theory implies a tight link between these different measures, considering their joint determination might also lead to more precise estimates. But until new research provides more refined and reliable estimates of potential GDP, we should likely heed Okun's (1962) warning that "meanwhile, the measure of potential must be used with care."

ACKNOWLEDGMENTS An earlier version of this paper was prepared for the second annual conference of the National Bank of Ukraine. Yuriy Gorodnichenko thanks the National Science Foundation for financial support. We are grateful to seminar participants at the Brookings Institution, the College of William & Mary, the University of Texas at Austin, the Central Bank of Chile, the Federal Reserve Bank of San Francisco, and the National Bureau of Economic Research's Summer Institute. We thank Laurence Ball, Olivier Blanchard, Aaron Flaaen, Serena Ng, Valerie Ramey, and James Stock for helpful comments, and Alexander Kopoin for sharing the OECD's data on potential output.

References

Andres, Javier, David Lopez-Salido, and Ed Nelson. 2005. "Sticky-Price Models and the Natural Rate Hypothesis." *Journal of Monetary Economics 52, no. 5:* 1025–53.

Anzoategui, Diego, Diego Comin, Mark Gertler, and Joseba Martinez. 2016. "Endogenous Technology Adoption and R&D as Sources of Business Cycle Persistence." NBER Working Paper 22005. Cambridge, Mass.: National Bureau of Economic Research.

Auerbach, Alan, and Yuriy Gorodnichenko. 2012a. "Fiscal Multipliers in Recession and Expansion." In *Fiscal Policy after the Financial Crisis*, edited by Alberto Alesina and Francesco Giavazzi. University of Chicago Press.

———. 2012b. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal–Economic Policy* 4, no. 2: 1–27.

Ball, Laurence. 2014. "Long-Term Damage from the Great Recession in OECD Countries." NBER Working Paper 20185. Cambridge, Mass.: National Bureau of Economic Research.

Barsky, Robert, and Eric Sims, 2011. "News Shocks and Business Cycles." *Journal of Monetary Economics* 58, no. 3: 273–89.

Baumeister, Christiane, and J. D. Hamilton. 2015. "Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks." Working paper. https://econweb.ucsd.edu/~jhamilto/bh2.pdf.

Beaudry, Paul, and Franck Portier. 2006. "Stock Prices, News and Economic Fluctuations." *American Economic Review* 96, no. 4: 1293–1307.

Benigno, Gianluca, and Luca Fornaro. 2018. "Stagnation Traps." *Review of Economic Studies* 85, no. 3: 1425–70.

Bernanke, Ben S., and Alan S. Blinder. 1992. "The Federal Funds Rate and the Channels of Monetary Transmission." *American Economic Review* 82, no. 4: 901–21.

Blagrave, Patrick, Roberto Garcia-Saltos, Douglas Laxton, and Fan Zhang. 2015. "A Simple Multivariate Filter for Estimating Potential Output." IMF Working Paper 2015–79. https://www.imf.org/external/pubs/ft/wp/2015/wp1579.pdf.

Blanchard, Olivier. 2017. "Should One Reject the Natural Rate Hypothesis?" Working paper, Peterson Institute of International Economics. https://piie.com/publications/working-papers/should-we-reject-natural-rate-hypothesis.

Blanchard, Olivier, and Daniel Leigh. 2013. "Growth Forecast Errors and Fiscal Multipliers." *American Economic Review* 103, no. 3: 117–20.

Blanchard, Olivier, and Danny Quah. 1989. "The Dynamic Effects of Demand and Supply Disturbances." *American Economic Review* 79, no. 4: 655–73.

Blanchard, Olivier, Guido Lorenzoni, and Jean-Paul L'Huillier. 2017. "Short-Run Effects of Lower Productivity Growth: A Twist on the Secular Stagnation Hypothesis." NBER Working Paper 23160. Cambridge, Mass.: National Bureau of Economic Research.

Braun, Steven. 1990. "Estimation of Current-Quarter Gross National Product by Pooling Preliminary Labor-Market Data." *Journal of Business and Economic Statistics* 8: 293–304.

Card, David, and Richard B. Freeman. 1993. "Introduction." In *Small Differences That Matter: Labor Markets and Income Maintenance in Canada and the United States*, edited by David Card and Richard B. Freeman. University of Chicago Press.

CBO (Congressional Budget Office). 2001. *CBO's Method for Estimating Potential Output: An Update*. August.

———. 2014. *Revisions to CBO's Projection of Potential Output Since 2007*. February.

Clark, Peter, 1979. "Potential GNP in the United States 1948–80." *Review of Income and Wealth* 25: 141–65.

———. 1982. "Okun's Law and Potential GNP." Board of Governors of the Federal Reserve System, October.

Cochrane, John. 1994. "Permanent and Transitory Components of GNP and Stock Prices." *Quarterly Journal of Economics* 109, no. 1: 241–65.

Coibion, Olivier, and Yuriy Gorodnichenko. 2012. "What Can Survey Forecasts Tell Us About Information Rigidities?" *Journal of Political Economy* 120, no. 1: 116–59.

———. 2015a. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review* 105, no. 8: 2644–78.

———. 2015b. "Is the Phillips Curve Alive and Well After All? Inflation Expectations and the Missing Disinflation." *American Economic Journal–Macroeconomics* 7, no. 1: 197–232.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia. 2017. "Innocent Bystanders? Monetary Policy and Inequality in the U.S." *Journal of Monetary Economics* 88: 70–89.

Coibion, Olivier, Yuriy Gorodnichenko, and Mauricio Ulate. 2019. "Is Inflation Just Around the Corner? The Phillips Curve and Global Inflationary Pressures." NBER Working Paper 25511. Cambridge, Mass.: National Bureau of Economic Research.

de Resende, Carlos. 2014. "An Assessment of IMF Medium-Term Forecasts of GDP Growth." Background Paper, Independent Evaluation Office of the International Monetary Fund.

Edge, Rochelle, and Jeremy Rudd. 2016. "Real-Time Properties of the Federal Reserve's Output Gap." *Review of Economics and Statistics* 98, no. 4: 785–91.

Erceg, Christopher, and Andrew Levin. 2014. "Labor Force Participation and Monetary Policy in the Wake of the Great Recession." *Journal of Money, Credit and Banking* 46, no. S2: 3–49.

Fatas, Antonio, and Lawrence Summers. 2018. "The Permanent Effects of Fiscal Consolidations." *Journal of International Economics* 112: 238–50.

Fernald, John G. 2007. "Trend Breaks, Long-Run Restrictions, and Contractionary Technology Improvements." *Journal of Monetary Economics* 54, no. 8: 2467–85.

———. 2012. "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity." Federal Reserve Bank of San Francisco Working Paper 2012–19.

Fernald, John G., James H. Stock, Robert E. Hall, and Mark W. Watson. 2017. "The Disappointing Recovery of Output after 2009." *Brookings Papers on Economic Activity*, Spring: 1–58.

Fleischman, Charles, and John Roberts. 2011. "From Many Series, One Cycle: Improved Estimates of the Business Cycle from a Multivariate Unobserved Components Model." Finance and Economics Discussion Series 2011–46. Washington: Divisions of Research and Statistics and Monetary Affairs of the Federal Reserve Board.

Friedman, Milton. 1957. *A Theory of the Consumption Function*. Princeton University Press.

Galí, Jordi. 1999. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review* 89, no. 1: 249–71.

Gonzalo, Jesus, and Serena Ng. 2001. "A Systematic Framework for Analyzing the Dynamic Effects of Permanent and Transitory Shocks." *Journal of Economic Dynamics and Control* 25, no. 10: 1527–46.

Gordon, Robert. 2014. "A New Method of Estimating Potential Real GDP Growth: Implications for the Labor Market and the Debt/GDP Ratio." NBER Working Paper 20423. Cambridge, Mass.: National Bureau of Economic Research.

Gordon, Robert, and Robert Krenn. 2010. "The End of the Great Depression 1939–41: Policy Contributions and Fiscal Multipliers." NBER Working Paper 16380. Cambridge, Mass.: National Bureau of Economic Research.

Gorodnichenko, Yuriy, and Matthew D. Shapiro. 2007. "Monetary Policy When Potential Output Is Uncertain: Understanding the Growth Gamble of the 1990s." *Journal of Monetary Economics* 54, no. 4: 1132–62.

Hauptmeier, Sebastian, Friedrich Heinemann, Marcus Kappler, Margit Kraus, Andreas Schrimpf, Hans-Michael Trautwein, and Qingwei Wang. 2009. *Projecting Potential Output: Methods and Problems*. ZEW Economic Study. Heidelberg: Physica-Verlag.

IEA (International Energy Agency). 2017. "World Indicators." IEA World Energy Statistics and Balances (database). http://dx.doi.org/10.1787/data-00514-en.

Jones, Charles I. 2016. "The Facts of Economic Growth." In *Handbook of Macroeconomics, Volume 2*, edited by John B. Taylor and Harald Uhlig. Amsterdam: Elsevier.

Keynes, John M. 1936. *A General Theory of Employment, Interest and Money*. New York: Palgrave Macmillan.

Kilian, Lutz. 2008. "Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy?" *Review of Economics and Statistics* 90, no. 2: 216–40.

———. 2009. "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market." *American Economic Review* 99, no. 3: 1053–69.

King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson. 1991. "Stochastic Trends and Economic Fluctuations." *American Economic Review* 81, no. 4: 819–40.

Knotek, Edward S., II. 2007. "How Useful Is Okun's Law?" *Federal Reserve Bank of Kansas City Economic Review* 92, no. 4: 73–103.

Kurmann, Andre, and Eric Sims. 2017. "Revisions in Utilization-Adjusted TFP and Robust Identification of News Shocks." NBER Working Paper No. 23142. Cambridge, Mass.: National Bureau of Economic Research.

Kuttner, Kenneth. 1994. "Estimating Potential Output as a Latent Variable." *Journal of Business and Economic Statistics* 12, no. 3: 361–68.

Mishkin, Frederic. 2007. "Estimating Potential Output." Speech given at Conference on Price Measurement for Monetary Policy, Federal Reserve Bank of Dallas, Dallas, May 24.

Moran, Patrick, and Albert Queralto. 2018. "Innovation, Productivity and Monetary Policy." *Journal of Monetary Economics* 93: 1–20.

OECD (Organization for Economic Cooperation and Development). 2012. "Medium and Long-Term Scenarios for Global Growth and Imbalances." In *OECD Economic Outlook, Volume 2012/1.* Paris: OECD Publishing.

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

Orphanides, Athanasios. 2001. "Monetary Policy Rules Based on Real-Time Data." *American Economic Review* 91, no. 4: 964–85.

———. 2003. "The Quest for Prosperity Without Inflation." *Journal of Monetary Economics* 50: 633–63.

———. 2004. "Monetary Policy Rules, Macroeconomic Stability, and Inflation: A View from the Trenches." *Journal of Money, Credit and Banking* 36, no. 2: 151–75.

Orphanides, Athanasios, and Simon van Norden. 2002. "The Unreliability of Output-Gap Estimates in Real Time." *Review of Economics and Statistics* 84, no. 4: 569–83.

Pagan, Adrian. 1984. "Econometric Issues in the Analysis of Regressions with Generated Regressors." *International Economic Review* 25, no. 1: 221–47.

Papell, David, and Ruxandra Prodan. 2012. "The Statistical Behavior of GDP after Financial Crises and Severe Recessions." *B.E. Journal of Macroeconomics* 12, no. 3: 1–31.

Ramey, Valerie. 2016. "Macroeconomic Shocks and Their Propagation." In *Handbook of Macroeconomics, Volume 2*, edited by John B. Taylor and Harald Uhlig. Amsterdam: Elsevier.

Ramey, Valerie A., and Sarah Zubairy. 2018. "Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data." *Journal of Political Economy* 126, no. 2: 850–901.

Romer, Christina, and David Romer. 2004. "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review* 94, no. 4: 1055–84.

———. 2010. “The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks.” *American Economic Review* 100, no. 3: 763–801.

Staiger, Douglas O., James H. Stock, and Mark W. Watson. 1997. “How Precise Are Estimates of the Natural Rate of Unemployment?” In *Reducing Inflation: Motivation and Strategy*, edited by Christina Romer and David Romer. University of Chicago Press.

Tarullo, Daniel K. 2017. “Monetary Policy Without a Working Theory of Inflation.” Brookings Hutchins Center Working Paper 33.

Wicksell, Knut. 1898. *Interest and Prices*. English translation, New York: Sentry Press, 1936.

World Bank. 2018. *Global Economic Prospects: Broad Based Upturn, but for How Long?*” Washington: World Bank. https://openknowledge.worldbank.org/handle/10986/28932.

Yellen, Janet L. 2017. “From Adding Accommodation to Scaling It Back.” Speech to the Board of Governors of the Federal Reserve System, Washington, March 3. https://www.federalreserve.gov/newsevents/speech/yellen20170303a.htm.

Comments and Discussion

COMMENT BY

SERENA NG This paper by Olivier Coibion, Yuriy Gorodnichenko, and Mauricio Ulate is motivated by uncertainty over the state of the economy due to diverse estimates of potential output. As documented by George Perry (1977), there were similar concerns after the 1974–75 recession. Interest in how to measure potential output is just as strong now as it was then.

The premise of this paper is that a reasonable estimate of potential output should have certain key properties, among which are dynamic responses to shocks. Using estimates of potential output and shocks collected from a variety of sources, Coibion and colleagues find that potential output tends to respond too slowly to permanent shocks and too much to transitory shocks. This finding is robust across measures of potential output for the United States and other countries. The authors attribute the problem to a strongly persistent component that gets embedded into the potential output estimates. Because of this, the authors warn against making inference about GDP from revisions in potential output. The paper then proposes a new measure of potential output that depends only on permanent shocks. The methodology yields an output gap ranging from 5 to 10 percent in 2017. Explanations are then given for why the output gap might differ from the unemployment rate.

I enjoyed reading this creative paper, and appreciate the immense amount of work put into synthesizing the data from so many sources. I particularly like the idea of using auxiliary information (here, shocks constructed elsewhere) to validate estimates of latent variables. In what follows, I first provide a framework to help understand why measures of potential output might respond to shocks in the manner documented in the paper. I then turn to some issues with the proposed methodology. I suggest that the maintained assumption of a linear trend is largely responsible for the large output gap of 5 to 10 percent, which is at odds with an unemployment rate that is near a historical low.

In what follows, I let Y_t be log GDP, Y_t^* be the level of potential output, and $\hat{Y}^*_{(t|t)}$ be a nowcast estimate of Y_t^* that can be obtained from the Congressional Budget Office, the Greenbook of the Federal Reserve Bank of Philadephia, or other sources. The proposed nowcast estimate is $\tilde{Y}^*_{(t|t)}$. I let $\hat{Y}_t^* = \hat{Y}^*_{(t|T)}$ be the full sample (smoothed) estimate. In this notation, the paper considers the regression

$$\Delta Y^*_{t|t} = a + \sum_{k=0}^{K} \phi_k \hat{\epsilon}_{t-k} + error \tag{1}$$

where $\hat{\epsilon}_t$ is one of the shock series collected from a variety of sources and $K = 8$. The base case uses the Greenbook's $\Delta\hat{Y}^*_{(t|t)}$ over the sample 1987:Q1–2016:Q4. With the responses of ΔY_t as a benchmark, the paper by Coibion and colleagues reports (1) insufficient sensitivity of $\Delta\hat{Y}^*_{(t|t)}$ to some permanent shocks and incorrect response to other permanent shocks; (2) excess sensitivity of $\Delta\hat{Y}^*_{(t|t)}$ to monetary policy and government spending shocks that should only have transitory effects; and (3) a pattern of dynamic responses that can be replicated using a one-sided Hodrick–Prescott (HP) filter with $\lambda = 500K$ as Y_t^*, leading to the suggestion that "there seems to be little value added in estimates of potential GDP relative to simple measures of statistical trends."

A quick remark on the third result. Even though the dynamic responses of $\hat{Y}^*_{(t|t)}$ are similar to those of an HP trend with $\lambda = 500K$, there is little in Coibion and colleagues' analysis to suggest that the level of such an HP trend is similar to the level of any of the $\hat{Y}^*_{(t|t)}$ that were carefully constructed. Furthermore, an HP trend is mean-squared optimal only if the cycle is white noise. As James Hamilton (2017) points out, $\lambda = 1600$ is already larger than the data-determined value. Thus, an HP trend with $\lambda = 500K$ cannot be seen as optimal in any meaningful sense.

UNDERSTANDING EXCESS AND INSUFFICIENT SENSITIVITY Results 1 and 2 hold for all five U.S. estimates of Y_t^* as well as for international data. It is thus useful to dig deeper into this. The shocks used in the analysis are themselves estimated, and their properties can in principle be questioned, but I abstract from this possibility. I suggest below that the first result is generic of filtering integration of order one [I(1)] processes, while the second result is symptomatic of a cyclical component that is strongly persistent.

Potential output plays a prominent role in policy work, but the variable is latent. Different estimates are obtained under different assumptions that are often not made explicit. To make sense of Coibion and colleagues' analysis, potential output must have a unit root (stochastic trend) component. Thus, let Y_t be the sum of a trend component τ_t and a stationary cyclical component c_t. The trend is itself the sum of two components: a deterministic trend d_t, a stochastic trend s_t:

$$Y_t = \tau_t + c_t \tag{2}$$

$$d_t + s_t + c_t. \tag{3}$$

I broadly define potential output as

$$Y_t^* = \mathbb{E}[\tau_t | \mathcal{R}] = \mathbb{E}[d_t + s_t | \mathcal{R}].$$

The challenge in identifying Y_t^* is that d_t, s_t, and c_t are latent, so we impose statistical and/or economic restrictions represented by $\mathcal{R}$. Different values of $\mathcal{R}$ lead to different estimates. As shown by Arthur Okun (1983), Y_t^* cannot generate inflationary pressure. The HP filter constrains ΔY_t^* to change slowly, while a production function approach requires Y_t^* to be consistent with full employment. As is seen below, Coibion and colleagues restrict the transitory shocks to have no long-run effect on Y_t^* and produce an estimate of s_t under the maintained assumption that d_t is a linear trend.

To understand results 1 and 2, suppose that

$$s_t = s_{t-1} + e_t^s \tag{4}$$

$$a(L)c_t = e_t^c. \tag{5}$$

I assume, for simplicity, that (e_t^s, e_t^c) are serially and mutually uncorrelated innovations to the trend and the cycle, respectively. For understanding results 1 and 2, there is also no loss in assuming that $d_t = a + gt$ is a linear trend, where the growth rate g is known. The assumption that s_t is a random walk is without loss of generality. However, it is important that c_t is stationary ergodic: $\alpha(z) = 1 - \phi_1 z - \phi_2 z^2 - \phi_p z^p \neq 0$ for $|z| \leq 1$,

In the simplest case, when $c_t = \alpha c_{t-1} + e_t^c$, ΔY_t can be represented by

$$\begin{aligned} \Delta Y_t &= \Delta d_t + \Delta s_t + \Delta c_t \\ &= g + \psi_y^s(L)e_t^s + \psi_y^c(L)e_t^c. \end{aligned} \tag{6}$$

Under my assumptions, the true long-run effect of a unit permanent shock e_t^s on ΔY_t is $\psi_y^s(1) = \psi_s^s(1) = 1$, while the long-run effect of a unit transitory shock on ΔY_t is $\psi_y^c(1) = 0$, because $\psi_y^c(L) = \psi_c^c(1) = \dfrac{1-L}{1-\alpha L}$ is zero, evaluated at $L = 1$.

Suppose in addition to Y_t, we observe e_t^s, e_t^c, d_t, and s_t. Note, first, that Δs_t is white noise and c_t is serially correlated with a first-order autocorrelation coefficient of α. My table 1 reports results for four values of α: 0, 0.25, 0.8, and 0.95—with $T = 200$. The sample autocorrelation coefficients when s_t and c_t are observed correctly reflect the dynamics of the two stochastic processes, as seen from the row denoted (ρ_s, ρ_c) in the table's fourth and fifth columns. Least squares regressions should recover the dynamic effects of e_{t-j}^s and e_{t-j}^c on Δs_t. Let $(\psi_{s,j}^s, \psi_{s,j}^c)$ denote these regression coefficients for $j = 1, \ldots, J$. The impact effect is given by $j = 0$ and the cumulative effect is captured by summing the coefficients from $j = 0$ to J. Because s_t is a random walk, one would expect $\psi_{s,0}^s = 1$ and $\psi_{s,j}^s = 0$ for $j > 1$. Furthermore, $\psi_{s,j}^c$ should be zero for all j because e_{t-j}^c should have no effect on s_t at any lag.[1] As shown in the fourth and fifth columns of table 1, the estimates of $\psi_{s,j}^s$ and $\psi_{s,j}^c$ have the values that we expect.

Now suppose we regress ΔY_t instead of Δs_t on the shocks and denote the coefficients by $(\psi_{y,j}^s, \psi_{y,j}^c)$. The second and third columns of my table 1 indicate that the cumulative effects are close to the true values of (1,0) only when α is small. At $\alpha = 0.95$, the estimates of (0.909,0.229) are biased. Though the downward bias of $\hat{\psi}_y^s$ is largely gone when $T = 2{,}000$, $\hat{\psi}_y^s$ remains biased at 0.146. This suggests, on one hand, that the problem is not just a finite sample issue, but also that $\hat{\psi}_y^s$ and $\hat{\psi}_y^c$ may not be reliable benchmarks because they are biased at precisely the parameter region of interest.

We do not observe Y_t^* or its components d_t and s_t. Applying a one-sided filter $H(L)$ to Y_t gives the $\hat{Y}_t^* = H(L)Y_t$

$$\Delta\hat{Y}_t^* = \Delta H(L)Y_t = H(1)g + H(L)e_t^s + \frac{H(L)(1-L)}{1-\alpha L}e_t^c$$

$$= g^* + \psi_*^s(L)e_t^s + \psi_*^c(L)e_t^c.$$

1. Coefficients with t statistics with less than 1 in absolute value are set to zero to get a more precise estimate of the long-horizon effect.

Table 1. Simulations to Illustrate Excess Sensitivity and Excess Smoothness[a]

DGP: $Y_t = d_t + s_t + c_t = \tau_t + c_t$

$\Delta s_t = e^s_t, \qquad e^s_t \sim N(0, .03)$

$(1 - \alpha L)c_t = e^c_t, \qquad e^c_t \sim N(0, .3), \qquad \mathrm{corr}(e^s_t, e^c_t) = 0.$

Regression: $Z_t = a + \psi^s_{Zj} e^s_{t-j} + \psi^c_{Zj} e^c_{t-j} + \text{error}.$

j	$\hat{\psi}^s_{y,j}$	$\hat{\psi}^c_{y,j}$	$\hat{\psi}^s_{s,j}$	$\hat{\psi}^c_{s,j}$	$\hat{\psi}^s_{MM,j}$	$\hat{\psi}^c_{MM,j}$	$\hat{\psi}^s_{HP,j}$	$\hat{\psi}^c_{HP,j}$
				$T = 200, \alpha = 0.0$				
0	1.003	0.999	1.000	–0.000	0.057	–0.001	0.103	0.101
1	–0.003	–1.000	0.000	0.001	0.042	–0.000	0.070	–0.022
2	–0.007	–0.000	–0.000	–0.000	0.044	–0.000	0.060	–0.010
0–12	0.009	0.001	0.000	0.001	0.043	–0.000	0.045	–0.005
0–40	1.013	–0.004	1.000	–0.003	0.381	–0.064	0.984	0.021
(ρ_s, ρ_c)			–0.005	–0.003	0.096	0.613	–0.015	0.422
				$T = 200, \alpha = 0.25$				
0	0.989	0.999	1.000	–0.001	0.059	–0.001	0.100	0.099
1	0.017	–0.750	–0.000	–0.000	0.043	–0.000	0.071	0.003
2	–0.025	–0.187	0.000	0.001	0.043	–0.001	0.058	–0.008
0–12	0.007	–0.011	0.000	0.000	0.043	–0.001	0.047	–0.007
0–40	0.933	–0.002	1.000	–0.003	0.398	–0.091	0.984	0.022
(ρ_s, ρ_c)			–0.002	0.240	0.123	0.691	0.058	0.548
				$T = 200, \alpha = 0.8$				
0	0.993	0.999	1.000	–0.001	0.057	0.008	0.102	0.103
1	–0.006	–0.201	0.000	–0.001	0.042	0.002	0.071	0.052
2	0.005	–0.161	0.000	–0.001	0.042	0.001	0.058	0.032
0–12	0.010	–0.103	–0.000	–0.000	0.040	0.000	0.047	0.010
0–40	1.057	0.014	1.000	–0.007	0.390	–0.289	0.985	0.086
(ρ_s, ρ_c)			–0.002	0.780	0.268	0.880	0.236	0.811
				$T = 200, \alpha = 0.95$				
0	0.992	1.000	1.000	–0.001	0.055	0.033	0.104	0.103
1	0.002	–0.049	–0.000	–0.001	0.041	0.025	0.066	0.068
2	–0.002	–0.044	0.000	0.001	0.041	0.024	0.056	0.055
0–12	0.006	–0.040	0.000	0.000	0.041	0.022	0.047	0.038
0–40	0.909	0.229	1.000	–0.004	0.313	–0.314	0.952	0.422
(ρ_s, ρ_c)			–0.006	0.929	0.391	0.938	0.308	0.912
				$T = 200, \alpha = 0.95$				
0	1.000	1.000	1.000	–0.000	0.049	0.032	0.042	0.043
1	0.001	–0.050	0.000	0.000	0.049	0.029	0.039	0.038
2	0.002	–0.048	0.000	–0.000	0.049	0.028	0.038	0.034
0–12	0.002	–0.043	0.000	0.000	0.049	0.025	0.035	0.029
0–40	1.011	0.146	1.000	–0.002	0.541	–0.240	0.973	0.359
(ρ_s, ρ_c)			0.001	0.948	0.754	0.949	0.611	0.931

Source: Author's calculations.

a. Results are based on the mean over 1,000 replications. The parameter $\psi^x_{z,j}$ is the effect of e^x_{t-j} on Z_t. Results are reported for j = 0, 1, and 2, as well as the sum of the coefficients up to lag 12 and 40. The (ρ_s, ρ_c) reports the first-order autocorrelation of variable Δs_t and c_t when they are observed, or the estimated trend and cycle when they are not.

I consider two choices of $H(L)$. The first is a 20-period, moving-mean filter (denoted MM in the sixth and seventh columns of my table 1). The second is the one-sided HP filter with $\lambda = 500K$ (denoted HP in the table's eighth and ninth columns). As seen from the first-order autocorrelation coefficient of $\Delta\hat{Y}_t^*$, the series exhibits correlation, even though Δs_t is white noise by design, and the gap estimate is more persistent than the true cycle. As seen from (ρ_s, ρ_c) for the last two columns of my table 1, $\Delta\hat{Y}_t^*$ is less persistent than Δs_t, while $\hat{c}_t$ is much more persistent than c_t. This is merely echoing the findings of Timothy Cogley and James Nason (1995), and many others, about the consequences of filtering.

What do we get when we regress $\Delta\hat{Y}_t^*$ instead of Δs_t on the shocks? Evidently, $\hat{\psi}_*^s(L)$ (with $* = MM$ or HP) is severely biased for the true value of 1, while the estimated effects of e_{t-j}^s on $\Delta\hat{Y}_t^*$ all differ from the value of zero. This arises because $H(L)$ spreads out the effect of the permanent shock over time. Result 1 documented in the paper by Coibion and colleagues—that $\Delta\hat{Y}_{t|t}^*$ reacts insufficiently to the permanent shock—is consistent with the simulation results. This bias is largely invariant to the dynamics of c_t.

In contrast, result 2 depends on the dynamics of c_t. Though the effect of the transitory shock e_t^c on the Δs_t is zero, the effect on ΔY_t^* is not. As seen from my table 1, the MM estimate of ΔY_t^* yields estimates of ψ_*^c that are similar in magnitude each period, a reflection of the constant weights in the MM filter. Though the bias at each lag is small, the bias in the cumulative effect is not. The bias in the HP estimates reflect the declining pattern of the HP weights. For both filters, the bias in the cumulative effect grows as α increase. When $\alpha = 0.95$, the cumulative effect of 40 lags is –0.314 for MM and 0.422 for HP. The bias is opposite in sign; hence, the choice of filter matters. This finding of excess sensitivity to transitory shocks is robust to changing alternative trend specifications as long as c_t has an autoregressive root local to unity.

The excess sensitivity result can be traced to the discontinuity of $\psi_*^c(L) = \dfrac{H(L)(1-L)}{1-\alpha L}$ at $\alpha = 1$. When α is far from 1, the long-run effect of e_t^c as measured by $\psi_*^c(1)$ is zero, because $1 - L = 0$ evaluated at $L = 1$. However, the term is of order 1 when $\alpha = 1$. When α is close to 1, there is a near-cancellation of $(1 - L)$ in the numerator, with $(1 - \alpha L)$ in the denominator, and e_t^c will appear as if it has permanent effects in finite samples. In other words, when c_t is highly persistent, it can be mistaken for s_t. The simulations in my table 1 bear this out; when c_t is highly persistent, transitory shocks have effects on $\Delta\hat{Y}_t^*$ that persist even after 40 periods.

One may ask if a cyclical component that is highly persistent is realistic. My own estimation of an unobserved components model (not reported) finds that the largest autoregressive root in c_t is bigger than 0.95. James Morley, Charles Nelson, and Eric Zivot (2003), and, more recently, Angelia Grant and Joshua Chan (2017) reported similar estimates for α. The autoregressive root in the unemployment series (more on this below) is also suggestive of a strongly persistent c_t. Now the half-life of a shock when $\alpha = 0.95$ is $\log(0.5) \div \log(0.95) = 14$ periods and increases to $\log(0.5) \div \log(0.98) = 34$ periods when $\alpha = 0.98$. Thus, though the observation by Coibion and colleagues that $\Delta\hat{Y}^*_{(t|t)}$ tends to respond to transitory shocks cannot be disputed, the finding that the responses are nonzero after eight quarters is not informative as to whether the long-run responses will be zero.

Coibion and colleagues have identified an interesting feature of many nowcast estimates of potential output that surprisingly has gone unnoticed. I conjecture that the finding will also hold for the smoothed estimates of potential output because the root problem is a cyclical component in GDP that is highly persistent, not data revisions.

THE PROPOSED ESTIMATE OF Y^*_T Coibion and colleagues' premise is that the $\hat{Y}^*_{(t|t)}$ measures are contaminated by persistent cyclical variations. But how to remove the "nearly permanent" cyclical component from these estimates? Bias-adjusting $\Delta\hat{Y}^*_{(t|t)}$ is difficult because we know how to estimate the purely permanent purely transitory quantities, but we are not very good at dealing with nearly permanent ones. Coibion and colleagues do not bias-adjust existing estimates but try something different. For $\Delta Y_t = g + \theta^s_t(L)e^s_t + \theta^c_t(L)e^c_t$, their idea is to define potential output growth as

$$\text{(7)} \qquad \begin{aligned} \Delta Y^*_{(t|t)} &= \mathbb{E}_t\,[\Delta Y_t | e^s_t, \ldots e^s_0, \ldots\,] \\ &= g + \theta^s_t(L)e^s_t \end{aligned}$$

where $e^s_t \ldots e^s_0$ are permanent shocks up to period t. Given real-time estimates of these shocks from a rolling-window application of the Blanchard-Quah (BQ) method, Coibion and colleagues then construct

$$\text{(8)} \qquad \Delta\tilde{Y}^*_{(t|t)} = g + \hat{\theta}^s_t(L)\hat{e}^s_{(t|t)}.$$

The proposed potential output $\tilde{Y}^*_{(t|t)}$ is the cumulative sum of $\Delta\tilde{Y}^*_{(t|t)}$, with $g = 0.031$, which is the average growth rate of Y over the sample period 1977–2007. This notion of trend output is similar in spirit to Beveridge and Nelson's (1981) trend. Both methods aim to produce an estimate of s_t assuming d_t is a linear trend. But Coibion and colleagues use the nowcast (instead of forecast) of output growth, their analysis is bivariate instead of univariate, and they take the extra step to make a permanent/transitory decomposition of the shocks.[2]

Estimation of the permanent shocks using the BQ method comes with some caveats. First, it depends on one variable (output) being I(1) and one variable (the unemployment rate) being I(0). But the unemployment rate has an AR(1) coefficient of 0.97. This is computed over the sample 1961:Q1–2016:Q4. Furthermore, the methodology depends on the choice of variables. Using capacity utilization in lieu of the unemployment rate, for example, will give a different estimate of potential output. Instead of two variables, Coibion and colleagues could have used a bigger vector autoregression, as is done by Anders Warne (1991) and by Jesús Gonzalo and me (2001), or they could have identified the stochastic trend directly—as is done by James Stock and Mark Watson (1988), Robert King and others (1991), and Gonzalo and Clive Granger (1995)—without going through the step of identifying the underlying shocks. A larger cause of concern is that all permanent/transitory decompositions implicitly or explicitly rely on estimates of the spectral density at frequency zero from the data.[3] If all carefully constructed $\hat{Y}^*_{(t|t)}$ series considered in Coibion and colleagues' paper have failed to isolate the pure trend component, one cannot be overly optimistic that the BQ methodology can succeed in doing so with 20 years of data. The standard errors around $\tilde{Y}^*_{(t|t)}$ must be unacceptably large.

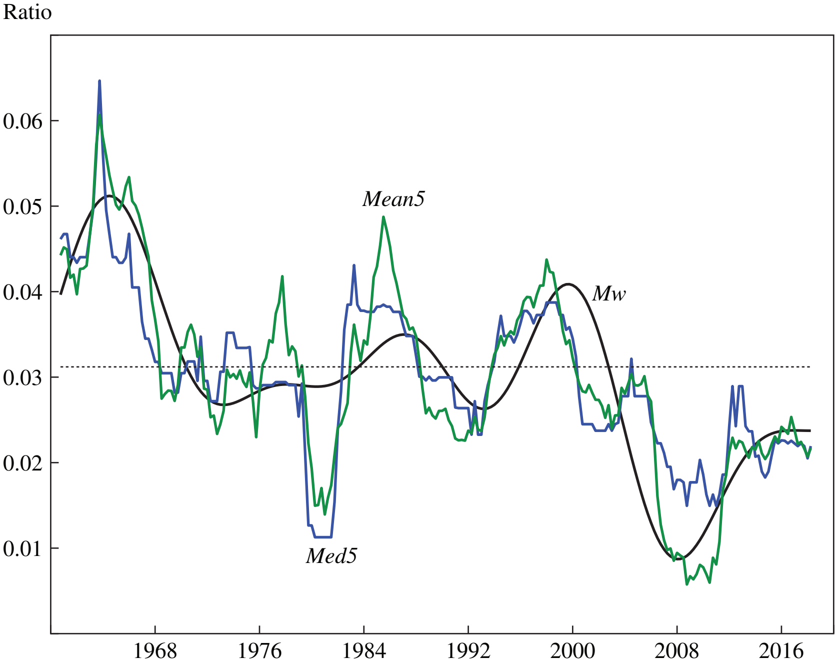
THE IMPORTANCE OF D_t Although having a measure of Y_t^* that is not affected in the long run by transitory shocks is desirable, the level of $Y_t^* = \mathbb{E}[d_t + s_t | \mathcal{R}]$ is of interest, not the counterfactual response of s_t to shocks. For this, the assumption on d_t becomes important. I will suggest that their implied output gap of 5 to 10 percent is due to the questionable assumption of the linear deterministic trend.

Even though Coibion and colleagues performed a rolling window estimation, $\tilde{Y}^*_{(t|t)}$ is still based on the assumption that g is a constant 0.031

2. Assuming that $d_t = \alpha + gt$ and s_t is a random walk, Beveridge and Nelson (1981) define $s_t^{BN} = \lim_{k\to\infty} \mathbb{E}_t (Y_{t+j} - kg) = Y_t + \lim_{k\to\infty} \sum_{k=1}^{\infty} \mathbb{E}_t (\Delta Y_{t+k} - g)$. Beveridge and Nelson's (stochastic) trend adjusts the current level Y_t by deviations of future growth from mean g.

3. In particular, the BQ and Galí (1999) methods essentially impose zero restrictions on the spectral density at frequency zero, while the Cochrane (1994) and Gonzalo and Ng's (2001) methods rely on cointegration arguments and still need some restriction on the spectral density at the zero frequency.

Figure 1. Smoothed GDP Growth, 1960–2020[a]



Source: Author's calculations.

a. The mw line is the low-frequency component of annualized GDP growth, as constructed by Mueller and Watson (2008). The med5 line is the five-year moving median, and the mean5 line is the five-year moving mean of the series.

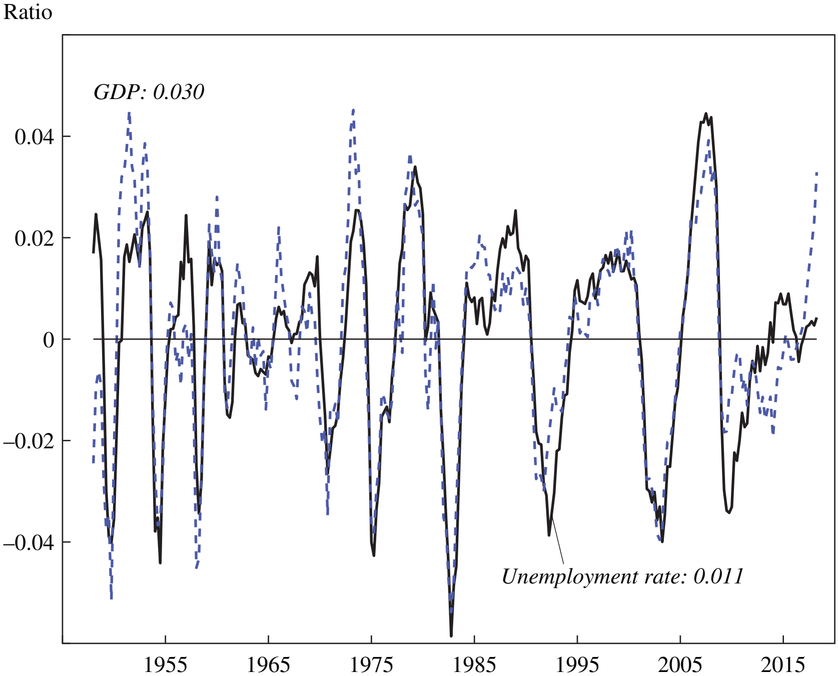
throughout. Over the sample 1948:Q1–2018:Q2 the residuals from linear detrending—that is, from regressing Y_t on a constant and a trend—have been –12 percent on average since 2012 and have been becoming more negative. The magnitude is in line with the –5 to –10 percent gap implied by $\tilde{Y}^*_{(t|t)}$. But a linear trend is monotone and cannot adapt to changes in demographics, technology, or any other structural aspects that have evolved over time. A quadratic trend that bends, for example, yields a gap of about –2 percent, and the second-order term is strongly statistically significant. The peak-to-peak method considered by Bradford De Long and Lawrence Summers (1988) gives a gap of about –1 percent. This is not to say that these methods are optimal, but rather that the linear trend is too rigid and leaves too much predictable variation unexplained to be desirable. My figure 1 shows the five-year moving mean and moving median of annualized GDP growth, along with the low-frequency component in the series estimated by the procedure of Ulrich Mueller and Watson (2008). It is then clear that the issue is not whether g should be 3.1 percent or 2.6 percent, but rather that it is assumed to be the same constant throughout. Because growth is well below the overall mean of 3.1 percent in the last 10 years, the constant growth assumption will overestimate Y_t^* during this period, giving a large output gap. Pierre Perron (1989) finds that misspecifying d_t can lead one to conclude that a unit root is present when the data are actually trend stationary. Misspecification of d_t will also affect the estimation of Y_t^*.

It is not a trivial task to disentangle s_t from c_t when c_t is strongly persistent, even if d_t is known. When d_t is itself of unknown form, as seen from my figure 1, the exercise becomes a formidable task. A different way to see the problem is that any s_t that is a unit root process is $Op(T)$. Any polynomial time trend is at least $O\left(T^{\frac{3}{2}}\right)$, so d_t dominates s_t when both are present. We cannot model s_t without first removing d_t. Given this difficulty, some may prefer to use the unemployment rate as a guide to the state of the economy. Coibion and colleagues provide compelling economic arguments for why the output gap is still a variable of interest. But from an econometric point of view, extracting a c_t from the unemployment rate (UR) is a more manageable exercise because it does not show a trend over time. As such, the d_t component of UR is just a constant and the only possible source of nonstationarity in UR is s_t. Identification of s_t is then much simpler, at least within the framework of unobserved components.

This, then, raises the question that perhaps the unobserved components model is asking too much of the data, and we should be content with being able to separate variations above and below certain frequencies. Mueller and Watson (2008) suggests a procedure (hereafter MW) that consists of projecting the series of interest on $K = 12$ cosine functions and taking the residuals as the cycle. After the low-frequency component is removed, the MW/UR gap still has an autoregressive root of about 0.92, similar to the one in the MW/output gap of about 0.89. My figure 2 plots the MW/output gap along with the MW/UR gap, but renormalized and centered to have the same mean and variance as the MW/output gap. We see that the two series match up remarkably closely over the last six decades. In 2016, which is the end of Coibion and colleagues' sample, there indeed appears to be more slackness in output than in the labor market, but much smaller than the 5 to 10 percent suggested by Coibion and colleagues. Both gaps suggest that the economy is near capacity in 2017.

In summary, Coibion and colleagues' results are consistent with a cyclical component in GDP that is strongly persistent. When the GDP data alone are uninformative about the trend component, using auxiliary information

Figure 2. The Unemployment Rate versus GDP: The Mueller–Watson Method, 1945–2020[a]



Source: Author's calculations.

a. The unemployment rate line is the cyclical component constructed by removing the low-frequency component of the unemployment rate. The GDP line is the cyclical component of log GDP. Both low-frequency and cyclical components are computed using the method of Mueller and Watson (2008).

to help with identification is potentially useful. These variables Z can be thought of "external" instruments. The question is how to use this information. Coibion and colleagues use shocks as Z and require that the sum of coefficients on the temporary shocks in equation 1 sum to zero. But their restriction only gives us a better estimate of s_t for a given d_t, while the level of Y_t^* is largely determined by d_t. The exercise is incomplete without a careful modeling of d_t.

REFERENCES FOR THE NG COMMENT

Beveridge, Stephen, and Charles R. Nelson. 1981. "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components, with Particular Attention to Measurement of the Business Cycle." *Journal of Monetary Economics* 7, no. 2: 151–74.

Cochrane, John H. 1994. "Permanent and Transitory Components of GNP and Stock Prices." *Quarterly Journal of Economics* 109, no. 1: 241–65.

Cogley, Timothy, and James M. Nason. 1995. "Effects of the Hodrick-Prescott Filter on Trend and Difference Stationary Time Series: Implications for Business Cycle Research." *Journal of Economic Dynamics and Control* 19, nos. 1–2: 253–78.

De Long, J. Bradford, and Lawrence H. Summers. 1988. "How Does Macroeconomic Policy Affect Output?" *Brookings Papers on Economic Activity*, no. 2: 433–80.

Galí, Jordi. 1999. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review* 89, no. 1: 249–71.

Gonzalo, Jesús, and Clive Granger. 1995. "Estimation of Common Long-Memory Components in Cointegrated Systems." *Journal of Business and Economic Statistics* 13, no. 1: 27–35.

Gonzalo, Jesús, and Serena Ng. 2001. "A Systematic Framework for Analyzing the Dynamic Effects of Permanent and Transitory Shocks." *Journal of Economic Dynamics and Control* 25, no. 10: 1527–46.

Grant, Angelia L., and Joshua C. C. Chan. 2017. "Reconciling Output Gaps: Unobserved Components Model and Hodrick-Prescott Filter." *Journal of Economic Dynamics & Control* 75: 114–21.

Hamilton, James D. 2017. "Why You Should Never Use the Hodrick-Prescott Filter." *Review of Economics and Statistics* 100, no. 5: 831–43.

King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson. 1991. "Stochastic Trends and Economic Fluctuations." *American Economic Review* 81, no. 4: 819–40.

Morley, James C., Charles R. Nelson, and Eric Zivot. 2003. "Why Are the Beveridge-Nelson and Unobserved-Components Decompositions of GDP So Different." *Review of Economics and Statistics* 85, no. 2: 235–43.

Mueller, Ulrich K., and Mark W. Watson. 2008. "Testing Models of Low-Frequency Variability." *Econometrica* 76, no. 5: 979–1016.

Okun, Arthur M. 1983. "Potential GNP: Its Measurement and Significance." In *Economics for Policymaking: Selected Essays of Arthur M. Okun*, edited by Joseph A. Pechman. MIT Press.

Perron, Pierre. 1989, "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis." *Econometrica* 57, no. 6: 1361–1401.

Perry, George L. 1977. "Potential Output and Productivity." *Brookings Papers on Economic Activity*, no. 1: 11–47.

Stock, James H., and Mark W. Watson. 1988. "Testing for Common Trends." *Journal of the American Statistical Association* 83, no. 404: 1097–1107.

Warne, Anders. 1993. "A Common Trends Model: Identification, Estimation, and Asymptotics." Working paper. Stockholm: Institute for Economic Studies.

COMMENT BY

VALERIE A. RAMEY This paper by Olivier Coibion, Yuriy Gorodnichenko, and Mauricio Ulate presents surprising new results showing that the leading real-time estimates of potential GDP for the United States and other industrialized countries react to temporary demand shocks. Potential GDP is intended to be an estimate of the maximum sustainable level of output that does not generate inflationary pressure. Because it is a supply-side concept, potential output should not react to demand shocks with temporary effects but should react fully to supply shocks with permanent effects. Coibion and colleagues present convincing evidence that none of the leading estimates of potential GDP satisfies this dichotomy.

Coibion and colleagues have three goals for their paper. Their first goal is to demonstrate that estimates of potential GDP by the various governmental and nongovernmental institutions in the U.S. and other industrialized countries overreact to shocks that have temporary effects on actual GDP and underreact to shocks that have permanent effects on actual GDP. The authors carefully construct real-time databases and use a variety of methods for estimating shocks to show convincingly that leading institutions, such as the Congressional Budget Office (CBO), revise their estimates of potential GDP in response to shocks that are easily identified, even in real time, as temporary. The authors estimate a variety of standard demand shocks, such as monetary and fiscal shocks, first showing that the impulse responses of actual GDP imply temporary effects and then showing that the estimates of potential GDP are revised in response to those shocks. They then estimate supply shocks, showing first that they have permanent effects on actual GDP and then that estimates of potential GDP are not revised sufficiently in response. Achieving this first goal constitutes two-thirds of the paper, and is its heart. These sections of the paper make a substantial contribution: the demonstration is very convincing, and the results are important because estimates of potential GDP are central to numerous quantitative models and are also important guides for policymakers. Perhaps one of the most surprising details in their findings is that estimates of potential GDP by the Federal Reserve's army of Ph.D. economists are virtually indistinguishable from a simple Hodrick–Prescott filter trend and that the Federal Reserve's own estimates of potential GDP are revised based on estimated monetary policy shocks. That is, the Federal Reserve's estimates of potential GDP behave as if monetary policy shocks have permanent supply-side effects, even though the impulse responses of actual GDP show no permanent effects of monetary policy shocks.

The paper's second goal is to explore alternative methods for estimating potential GDP that overcome the problems highlighted in the authors' demonstration. Their main suggested alternative is Olivier Blanchard and Danny Quah's (1989) decomposition of GDP shocks into demand and supply shocks using long-run restrictions, known as the BQ method. Coibion and colleagues show that this measure of potential GDP does not suffer from the same weaknesses as standard measures documented in the earlier sections of the paper. In addition, they explore a variety of other methods based on economic-theory with either alternative long-run restrictions based on theory or Phillips curves.

Finally, Coibion and colleagues' third goal is the production of an alternative measure of the current output gap. Using their implementation of the BQ method, they offer an alternative estimate of current potential GDP and conclude that actual GDP was still more than 5 percent below potential GDP in 2017.

I believe that Coibion and colleagues are very successful in achieving their first goal. Their careful demonstration of the weaknesses of current methods makes it clear that estimates of potential GDP can be improved. Regarding their second goal, their explorations of alternative methods are very promising. I believe that their choice of alternatives is very good. However, as I make clear below, there are remaining challenges with the implementation of their preferred alternative, so more work needs to be done. I demonstrate that key assumptions in their implementation lead to their implausible conclusion that current GDP is significantly below potential GDP. As a result, I do not think their estimates are ready for use by policymakers.

THE PAPER'S ALTERNATIVE METHODS FOR ESTIMATING POTENTIAL GDP To address the weaknesses of the standard estimates of potential GDP, Coibion and colleagues explore alternative methods for estimating potential GDP that can distinguish between shocks that have temporary versus permanent effects on actual output. The main alternative method they explore is the BQ decomposition method. This method uses a bivariate time series model with real GDP and the unemployment rate, and it identifies supply shocks as those shocks that have long-run effects on GDP and demand shocks as all other shocks that have temporary effects. Even if one does not agree with BQ's supply shock–versus–demand shock dichotomy, their method is still useful for separating out temporary from permanent shocks to GDP, which is the key to improving estimates of potential GDP.

Coibion and colleagues also explore other alternatives. For example, they use Jordi Galí's (1999) long-run restriction to identify permanent shocks to technology; John Cochrane's (1994) permanent income hypothesis-motivated method for using the behavior of consumption to identify permanent shocks to GDP; and a Phillips curve model to infer potential GDP from inflation dynamics. The authors' implementation of all these methods implies much larger current output gaps—that is, actual GDP is farther below potential GDP than those implied by the CBO's estimates and others.

I focus on Coibion and colleagues' implementation of the BQ method because that is their favored method, and that method actually gives a more conservative estimate of the gap relative to their other alternatives. Nevertheless, the authors' particular implementation of the BQ method implies a large gap. Their estimate of potential GDP leads them to conclude that "the gap between potential and actual output in the U.S. increased by about 5 log percentage points between 2007:Q1 (when the gap was likely close to zero) and 2017:Q1, leaving ample room for policymakers to close this gap through demand-side policies if they chose to do so." Thus, their estimates can be seen as an encouragement for policymakers to undertake more demand-side stimuli, even when the unemployment rate is below 4 percent.

IMPLICATIONS OF THE AUTHORS' POTENTIAL GDP ESTIMATES I now demonstrate that Coibion and colleagues' alternative estimates of potential, while avoiding the weaknesses they highlighted for the standard estimates, have a number of implications ranging from questionable to implausible. I argue, however, that the problem is that their estimates are based on questionable auxiliary identifying assumptions that are relatively easy to fix.

Implication 1: Coibion and colleagues' estimates of potential GDP decline as much as the CBO's estimates after the Great Recession. One of Coibion and colleagues' main critiques of the CBO revisions of potential GDP is that they lowered them too much from 2007 to 2017, in response to cyclical fluctuations. Figure 1 of their paper shows how the CBO's estimates of potential GDP changed from 2007 to 2017, and figure 13 shows how their own BQ estimates changed in real time from 2007 and 2017. Consider the revision for the estimate of potential output at the end of their sample, 2016:Q4. Using their data and programs, I calculated that the CBO revised down its estimate of potential GDP in 2016:Q4 by about 0.12 log points, whereas Coibion and colleagues' BQ estimate was revised down by about 0.11 log points over the same period. Thus, both methods lead to the same downward revision in potential GDP. If we believe that Coibion and colleagues' method is accurately capturing only permanent shocks, then their method *validates* the CBO revisions.

Implication 2: The implied natural rate of unemployment is implausibly low. We can combine Coibion and colleagues' estimate of the output gap with Okun's law to calculate the implied natural rate of unemployment. In their paper, Coibion and colleagues conduct this exercise in subsection IV.C. However, they use the older historical estimates of –3 for the parameter on the unemployment gap term rather than the more up-to-date estimates of –2 (Ball, Leigh, and Loungani 2017). Furthermore, they use their estimate of the output gap in 2016:Q4. Because the unemployment rate has fallen so much since then, adding more recent data is instructive.

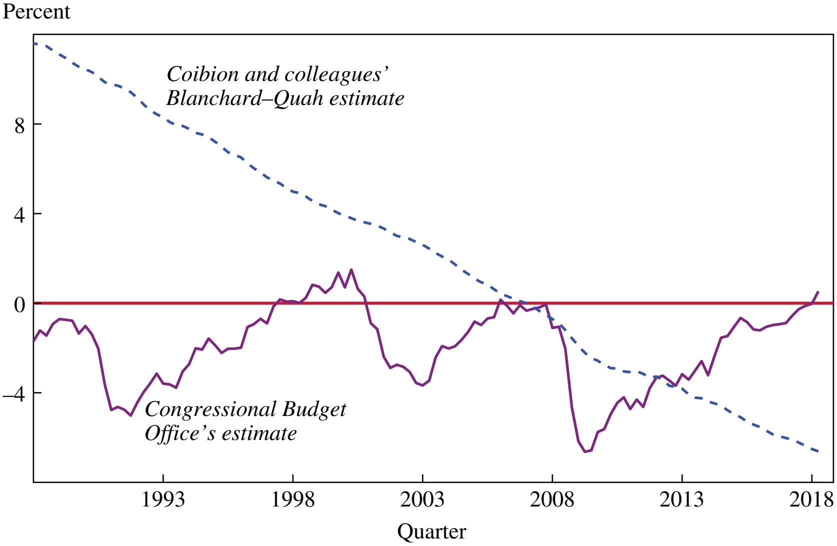
Thus, I update Coibion and colleagues' BQ estimates through 2018:Q2, using their same programs and the same rolling window over the previous 30 years. I find that actual output is about 6.6 percent below their estimate of potential GDP in 2018:Q2. Thus, using their method, I find that actual GDP is farther below potential GDP in 2018:Q2 than it was in 2016:Q4.

The unemployment rate in 2018:Q2 was 3.9 percent. Using Okun's law with a modern unemployment gap coefficient of –2 implies that the natural rate of unemployment in 2018:Q2 was about 0.6 percent. This unemployment rate is below any level ever achieved in the United States, including World War II, and is completely implausible.

Coibion and colleagues argue, however, that the usual Okun's law relationship no longer applies because the employment-to-population ratio in the U.S. fell so much during the Great Recession. Though this is an intriguing possibility, I show below that there is a much simpler explanation for why they estimate such a large output gap and implied low natural rate of unemployment: one of their auxiliary identifying assumptions leads potential GDP to have a significantly higher growth rate than actual GDP in the long run.

Implication 3: Coibion and colleagues' implied output gap has a strong upward trend. As mentioned briefly in discussing the last point, Coibion and colleagues' method for estimating potential GDP implies a bigger output gap in 2018 than at the end of 2016, which seems odd given the fast pace of growth of the U.S. economy and the significant decline in the unemployment rate. This feature led me to inspect Coibion and colleagues' implied output gap for the last 30 years more closely, because they use 30-year rolling regressions to counter possible breaks in trends. In my figure 1, I show the output gap estimated by the CBO and by Coibion and colleagues, where the gap is defined as log actual output minus log potential output so that the gap should be negative at the end of a recession. The CBO's gap behaves as expected, varying cyclically but with no trend. In contrast, the dominant feature of the Coibion and colleagues' implied gap

Figure 1. Estimated Output Gaps, $Y - Y^{*}$, 1988–2018[a]



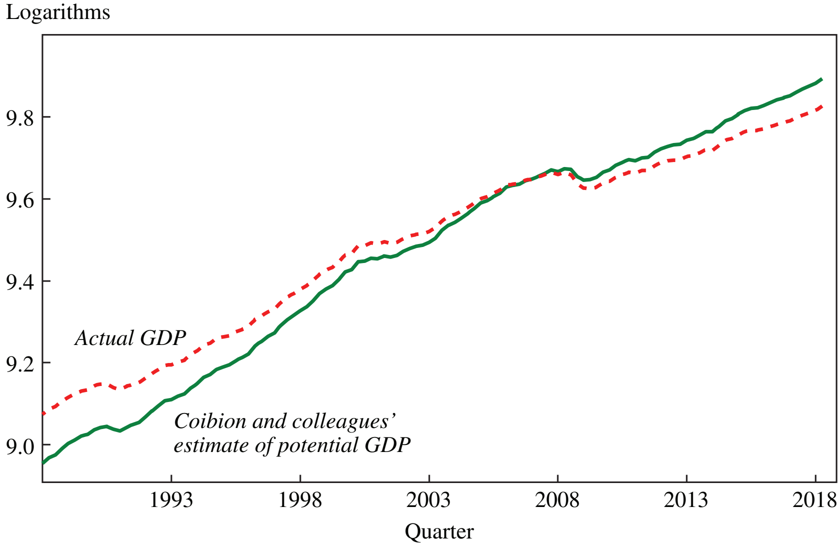
Sources: Author's estimates, using programs from Coibion and colleagues' paper and updated data from the FRED database of the Federal Reserve Bank of Saint Louis.

a. Y = log actual GDP; Y^* = log potential GDP.

is a strong downward trend—the estimated gap declines at a rate of about 0.6 percent per year. According to their estimates, the output gap was very positive in 1988, implying that *actual* output was almost 12 percent above *potential*. However, over time, this gap has narrowed and has become negative. According to the authors' estimates, the output gap is wider now, at about –6.6 percent, than it was at the end of the Great Recession, when it was about –2.2 percent.

This result is a direct consequence of Coibion and colleagues' estimated potential GDP having a much higher trend than actual GDP over the last 30 years. My figure 2 shows the path of both series. Even in the second half of the 1990s, when the growth of total factor productivity surged because of the information technology revolution, they estimate that actual GDP was significantly above potential GDP. The two series cross in 2007, and then the gap becomes negative and widens over time because their estimate of potential GDP grows more quickly than actual GDP. The next section explains which of the assumptions made by Coibion and colleagues lead to this implausible behavior.

Figure 2. Actual GDP versus Potential GDP, as Estimated by Coibion and Colleagues, 1988–2018



Sources: Author's estimates, using programs from Coibion and colleagues' paper and updated data from the FRED database of the Federal Reserve Bank of Saint Louis.

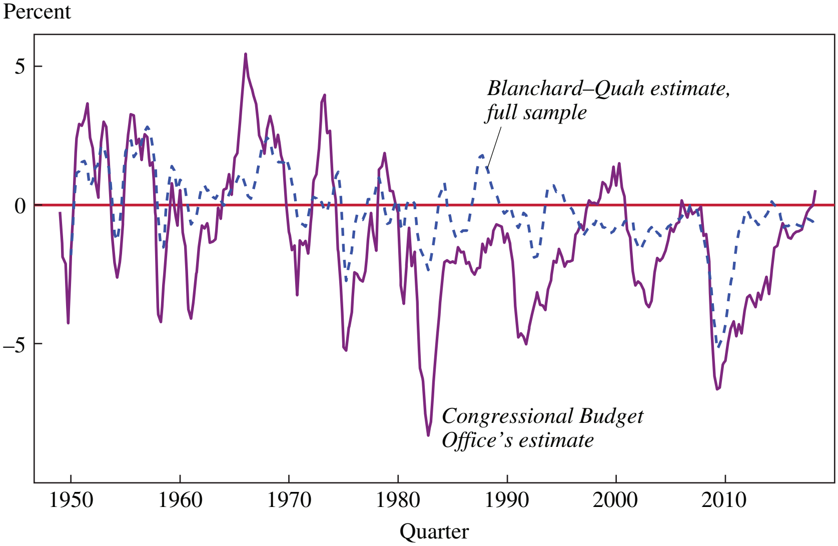
THE BQ METHOD IS NOT ENOUGH TO IDENTIFY POTENTIAL GDP Coibion and colleagues recognize that in order to implement the BQ method to derive a path of potential GDP, they must take a stand on the long-run growth rate of the economy. To see the identification problem, consider the intuitive equation they offer in subsection IV.A of their paper:

$$\Delta \log Y_t = g + \Delta \log Y_t^P + \Delta \log Y_t^C,$$

where $\Delta log Y_t$ is the growth rate of actual GDP, g is the long-run growth rate of GDP, $\Delta log Y_t^P$ is the growth rate of output due to permanent shocks, and $\Delta log Y_t^C$ is the growth rate of output due to temporary shocks. The BQ method assumes that permanent shocks can permanently affect the *level* of GDP, but not the *growth rate* of GDP. Therefore, the BQ method identifies only *deviations* from a long-run path; hence, neither the slope (g) of this path nor the intercept is identified.

Thus, Coibion and colleagues are forced to make two additional assumptions to identify the path. To identify the *slope* of the path, they assume a value of g of 3.1 percent, which equals both the average growth rate of

Figure 3. Estimated Output Gaps, $Y - Y^*$, 1950–2020[a]



Sources: Author's estimates, using programs from Coibion and colleagues' paper and updated data from the FRED database of the Federal Reserve Bank of Saint Louis.

a. The CBO's estimate versus the Blanchard–Quah method's estimate on the full sample.

real GDP from 1977 to 2007 and for the entire post–World War II period. To identify the *intercept* of the path, they assume that potential GDP was equal to actual GDP in 2007:Q1. Also, the CBO's estimated gap is then only about –0.3 percent, so this assumption is close to the CBO's estimates. However, as my figure 2 shows, the slope estimate for g leads the authors' estimate of potential GDP to grow much faster than actual GDP from 1988 to 2018. It is this divergence in growth rates that leads directly to their estimate that output is currently 6.6 percent below potential GDP.

The problem of different growth rates for actual and potential GDP would not occur if g were set equal to the actual growth rate of GDP over the sample used in the estimation. To demonstrate this, I updated the authors' data and reestimated their BQ model back to 1948 and created output gap estimates. These are shown in my figure 3, along with the CBO's estimates. As the figure shows, there is no longer a trend in the gap estimate. However, the two estimates do not move in lockstep. The correlation between the CBO's gap estimate and the BQ gap estimate is about 0.5, suggesting that much could be learned from the differences in the implied gaps.

CONCLUSION Overall, this is an important paper that effectively demonstrates that standard measures of potential GDP overreact to temporary shocks and underreact to permanent shocks. It makes a convincing argument that we can do better, even in real time. The alternative methods explored are promising, but the methods still need work, so any implied gap estimates are "not yet ready for prime time." For now, I think I will stick with the CBO's estimate of the gap, which indicates no slackness in the U.S. economy.

REFERENCES FOR THE RAMEY COMMENT

Ball, Laurence, Daniel Leigh, and Prakash Loungani. 2017. "Okun's Law: Fit at 50?" *Journal of Money, Credit and Banking* 49, no. 7: 1413–41.

Blanchard, Olivier J., and Danny Quah. 1989. "The Dynamic Effects of Aggregate Demand and Supply Disturbances." *American Economic Review* 79, no. 4: 655–73.

Cochrane, John H. 1994. "Permanent and Transitory Components of GNP and Stock Prices." *Quarterly Journal of Economics* 109, no. 1: 241–65.

Galí, Jordi. 1999. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review* 89, no. 1: 249–71.

GENERAL DISCUSSION James Stock began by drawing a firmament analogy, wondering if among the stars in the firmament, potential output—Y^*—had any contributions beyond the natural rate of unemployment—U^*. He postulated that in principle, the answer could be yes, because potential output can incorporate capital accumulation, total factor productivity growth, changes in underlying population growth, and changes in the labor force participation rate. This can provide additional information and help explain measures of slackness in the economy and, therefore, thinking about monetary and fiscal policy. However, each of these additional factors has many problems in practice. He acknowledged that perhaps it is plausible to forecast population growth or put aside immigration issues, but there are still ongoing challenges in understanding the labor force participation rate and total factor productivity growth. Although, in principle, it might be possible to get these things right—such as determining the underlying growth rate, and thus making measures of potential output more informative than the natural rate of unemployment—whether this can be pulled off in practice remains doubtful.

Stock stated that based on evidence presented by the authors and other evidence that he had seen, it was not clear to him if the additional challenges of moving to potential output from the natural rate of unemployment are worth it. It was also not clear to him if much more can be done than just estimating an output (Y) gap as an unemployment (U) gap times a rolling Okun's law coefficient (or something along those lines), and if there is any value in that. A plot of the output gap estimated by the Congressional Budget Office (CBO) against its estimation of the unemployment gap reveals that the two gaps are almost same, with a slight time variation in the Okun's law coefficient between the two plots. Therefore, in practice, the CBO's methodology states that it is tough to learn anything about an output gap that is not already observed in an unemployment gap, although this fact remains buried in the methodology. Stock thought that while the unemployment gap is challenging to measure, the scope for challenge is a little bit less than that for measuring the output gap. Going back to his firmament analogy, he concluded that despite the potential for learning more by looking at an output gap and potential output, it made sense for him to pull that star, Y^*, out of the firmament and to continue focusing on the natural rate of unemployment, U^*.

Jonathan Pingle asked the authors and commenter Valerie Ramey if they thought that a greater consideration of the 2005–7 period might be useful. Considering the paper's analysis of growth rate shocks, Pingle noted that the stepping off point was that the output gap in the last expansion barely seemed to close, although that is a little obscured by indexing in many of the estimates. However, the subsequent level of the gap matters a lot for policy. Pingle observed that in the 2005–7 period, inflation was running above the central bank target, despite globalization pressures, and there were imbalances in the economy, including overactivity in housing, a sector that is sensitive to the interest rate. This would imply an output gap that was more than just closed, or is inconsistent with an appropriate equilibrium target for monetary policy. He asked whether the level and stepping off point deserve more consideration, or whether a level of potential output that is too high is simply being carried forward.

Olivier Blanchard had two comments. The first was about the Blanchard–Quah approach, which is mentioned in the paper and the discussions. There are two conceptually separate steps to this approach. The first is a statistical decomposition of output between the part due to shocks that have a permanent effect on output, and shocks that only have a temporary effect. The second is the reference to the shocks with permanent effects as "supply shocks," and to the shocks with temporary effects as "demand shocks." The first step is simply data description and should be uncontroversial. The second step is controversial. One can buy the first step and construct a series for output due to shocks with permanent effects, and call it potential output, without accepting the second step.

Blanchard, drawing on his experience having seen the construction of potential output in various institutions, suggested that it often suffered from two problems. The first is intellectual laziness and the ease of using a simple statistical method, such as the Hodrick–Prescott (HP) filter, without thinking hard about the implicit assumptions behind it. The second includes political factors. Looking at retrospective revisions of the output gap in Greece is revealing: In real time, the assessment was that there was not much of an output gap. However, the output gap at the start of the crisis is now viewed as having been a large positive number, which makes the fall look less bad than it would otherwise be viewed as being. Blanchard suspected that though this is much less relevant for the CBO, it is nonetheless worth considering.

Turning to the question of what to do about measuring potential output and the output gap, Blanchard thought that economists need to look for signals of whether the economy is overheating or underheating. Inflation is far from a perfect signal, but it is the most natural one and the first one that should be considered. Blanchard did not think there was enough consideration of inflation by the authors, other than mentions of the Phillips curve. If economists really think that the inflation signal is becoming worse, which many do, then it is important to look at many dimensions of the labor market—such as the degree of labor force participation relative to a reasonable trend, the ratio of vacancies to unemployment, and the degree of involuntary part-time work—and then to use all these variables, together with inflation, to get to the natural rate of unemployment.

John Haltiwanger was struck by how few data go into producing potential output and real-time output, and he thought they are somewhat related. Starting with real output, he noted that statistical agencies continue to use very crude methods for their benchmark revisions, both methodologically and in timing. For example, the "birth-death model" used for the benchmark revisions from the payroll survey is really outdated in many ways and has always been crude. The Bureau of Labor Statistics is sitting on top of administrative data and could do a much better job. Haltiwanger acknowledged that the bureau is strapped for resources, so he did not want to pick on it, but that the general question is whether much more progress could be made in solving some of these problems with real-time data.

Moving on to potential output, Haltiwanger observed that there has been a lot of thinking about changes in potential output, building up from microeconomic evidence. A recent Jackson Hole conference was about these topics, such as changing market structure, concentration, and the changing role of start-ups.[1] Many classes of models suggest that the way to understand this is via heterogeneous firm and worker models and associated data. Economists are getting closer to real-time data on these factors. For example, the census has published a real-time business formation index that could be used in such contexts. Haltiwanger concluded that it might be possible to make progress on some problems by paying attention to the push toward heterogeneous agent models and the data that go along with them.

Eswar Prasad observed that one of the crucial issues in the literature was figuring out the right benchmark for evaluating different measures of output gaps. Much of the paper focuses on defining statistical benchmarks, which are very important. However, building on earlier discussions, Prasad argued that economic benchmarks are potentially far more important because of their implications for inflation or variables that the output gap may eventually affect. For instance, taking the Blanchard–Quah approach, which Prasad described as formidable, one could use the same model with inflation instead of the unemployment rate, because the identification restrictions would work very similarly if one made assumptions about how supply and demand shocks affect inflation. The right approach would be to slim down the number of economically meaningful variables and add more variables. Trying to infer what is happening with the output gap by looking at as many indicators as economic models might suggest would have some relationship with the output gap.

Prasad also stated that the univariate filter seemed to work very well. He recalled that back in 1991, when he was at the International Monetary Fund and was responsible for calculating measures of potential output, the HP filter was new and fresh, and so it seemed appropriate. However, even then there were concerns about whether the filter was too sensitive to observations

1. "Changing Market Structures and Implications for Monetary Policy," Conference, Jackson Hole, Wyo., August 23–25, 2018, https://www.kansascityfed.org/publications/research/escp/symposiums/escp-2018.

toward the end of the sample. The problem would be exacerbated by increasing the smoothing parameter, which is very sensitive to what happens at the end of the sample. Prasad recommended considering alternatives, like the King–Baxter band-pass filter, that have slightly better end-of-sample properties, and measure univariate filters against those slightly more robust alternatives, which are not as sensitive to end-of-sample problems.

Steven Braun stated that the one thing that he would have changed about the paper would be not using GDP, but instead using gross domestic output (GDO), which is the average of GDP and gross domestic income (GDI), because GDO has a higher correlation with the unemployment rate than either GDP or GDI does individually. Braun recalled having observed this in a past Brookings paper by Jeremy Nalewaik, and thought that that advice had been neglected.[2] Including Okun's law in the list of ways to calculate potential GDP given by the authors is the second thing that Braun would have changed. He noted that Okun's law, which is much simpler than the production function, is the first item that he would have listed. Yet even while using the production function, Okun's law comes in through the back door because of the adjustment from the actual labor force to the potential labor force. In addition, the Phillips curve must also be used, because a natural rate of unemployment is required to use Okun's law. Braun observed that the basic problem is that the Phillips curve has stopped working. An estimation of the Phillips curve that is restricted to the past 25 years shows that a zone of two-sigma uncertainty now includes plus or minus infinity. He concluded that it is difficult or impossible to estimate potential output without a natural rate of unemployment.

Robert Hall agreed with James Stock and reiterated that labor is the most important input to the economy. A measure like the output gap, which describes slackness or the lack of slackness, should be mapped into the labor market. Put differently, the economy is at potential when there is full employment. Today, every measure of the labor market screams tight, with no exception. Labor force participation has been considered an exception by some; however, those who have looked carefully at participation, including Hall himself, have concluded that there was a steep decline in participation that has not been erased by the restoration of full

2. Jeremy J. Nalewaik, "The Income- and Expenditure-Side Estimates of U.S. Output Growth," *Brookings Papers on Economic Activity*, Spring 2010, https://www.brookings.edu/bpea-articles/the-income-and-expenditure-side-estimates-of-u-s-output-growth/.

employment.[3] It would therefore be a mistake to incorporate changes in participation into a measure of labor market tightness. Hall noted that the labor market is as tight as it has ever been since the Current Population Survey was created under its present name in 1947–48, and all the reasonable measures, including those from the employer side, from the Job Openings and Labor Turnover Survey (the average duration of vacancies), are at an all-time high. All suggestions that unemployment is a bad measure of tightness have faded from influence, as has the notable persistence of long-term unemployment. Since 1948, there has been no trend in unemployment; it is a remarkably stable indicator, and this is reflected in the stability of Okun's law. Hall concluded that the measures of the output gap that track unemployment very closely are right, and any paper that says otherwise should be questioned.

Athanasios Orphanides applauded the paper, as he recalled earlier panels with Arthur Okun, Bob Hall, and George Perry presenting work on exactly the question that the paper tried to answer.[4] He noted the difficulty in identifying temporary effects as distinct from permanent effects, and he wondered how this translated into estimates of potential output and the corresponding implications for policy. The difficulty of separating temporary and permanent shocks is evidence for the need to identify robust ways of formulating countercyclical policy, monetary policy, and fiscal policy. He thought that the CBO is doing a very good job of this, considering the difficulties. In the case of monetary policy, economists have been making progress in recognizing that output gaps are mismeasured by downplaying the role of output gap measurement and taking more signals from inflation and inflation expectations. Employment gaps are useful; however, their measurement also has the issue of trying to evaluate the natural rate of unemployment. Although monetary policy has drawn these policy conclusions, the next item on the research agenda is finding out what advice can be drawn for fiscal policy. Orphanides wondered how uncertainty about long-term estimates of potential output can be incorporated into fiscal projections, taking into account the sensitivity from one-sided political pressures. Everybody is happy to raise estimates of potential, and using these

3. John G. Fernald, James H. Stock, Robert E. Hall, and Mark W. Watson, "The Disappointing Recovery of Output after 2009," *Brookings Papers on Economic Activity*, Spring 2017, 1–58, https://www.brookings.edu/wp-content/uploads/2017/08/fernaldtextsp17bpea.pdf.

4. George L. Perry, "Labor Force Structure, Potential Output, and Productivity," *Brookings Papers on Economic Activity*, no. 3, 1971, https://www.brookings.edu/bpea-articles/labor-force-structure-potential-output-and-productivity/.

estimates for policy. Conversely, however, everybody is unhappy when the estimates of potential output growth are reduced.

Robert Gordon agreed with James Stock and commented on how there had been two papers over two days—one on monetary policy,[5] and one on potential output,[6] both of which did not mention the unemployment rate. He defined potential output as a situation in which inflation is neither accelerating nor decelerating, which is exactly the same as the definition of the natural rate of unemployment. Therefore, by definition, the output gap would be zero when the unemployment gap is zero. Gordon reiterated Valerie Ramey's point, made while giving her comment, that the paper's conclusion that the output gap is currently 6 to 10 percent is implausible.

Regarding the questions of estimating potential output and the CBO's method of doing it implicitly, Gordon thought that James Stock had come close to the answer through his suggestion of aligning the output gap with the unemployment gap. A Kalman filter can be applied that extracts anything that is correlated with the unemployment gap from the cycles in output, using the unemployment gap as information. The result is a series of potential output data that is much more stable in comparison with that generated using the HP filter. This series does not respond to the decline in actual output during the 1981–82 recession and behaves similarly in the years 2007–9. It slows down radically after 2009, not in response to the demand decline but because of the underlying decline in the growth rate of productivity and the decline in labor force participation. Therefore, potential output backed out from the unemployment gap is radically slow growing and suggests a zero output gap in the current economy. Gordon also responded to Steve Braun's comments. He agreed with Braun that using the GDO in studying the response of productivity and output per hour produces very stable results. Regarding Braun's comments on the range of uncertainty of infinity and the disappearance of the Phillips curve, Gordon recommended waiting. He noted that the core personal consumption expenditure inflation had risen from about 1.4 to 2.0 percent in the previous year and that the Federal Reserve had forecasted continued 2.1 percent inflation over the next two years without any upward movement in inflation. One

5. Emmanuel Farhi and François Gourio, "Accounting for Macroeconomic and Finance Trends: Market Power, Intangibles, and Risk Premia," in this issue of *Brookings Papers on Economic Activity*.

6. Olivier Coibion, Yuriy Gorodnichenko, and Mauricio Ulate, "The Cyclical Sensitivity in Estimates of Potential Output," in this issue of *Brookings Papers on Economic Activity*.

would need to wait for two years to know whether the Phillips curve has truly disappeared.

Steven Davis agreed with Robert Hall's comments about the natural rate of unemployment but saw the current unemployment picture as murky. In particular, the frictional rate of unemployment is lower today than it was 10, 20, or 30 years ago—for two reasons. First, the labor force has aged considerably since the 1980s. Older workers have fewer short-term unemployment spells, which leads to a lower frictional rate of unemployment. Second, the frictional rate of unemployment has also fallen because of a trend decline in business volatility and job reallocation rates since the 1980s. Davis then referred to Valerie Ramey's estimate of a 3.5 percent natural rate of unemployment using the Blanchard–Quah methodology. Although Davis does not see 3.5 percent as his point estimate for the natural rate of unemployment, he does not find it outside the realm of plausibility, given the forces driving the decline in frictional unemployment. He concluded by noting that there is a fair degree of uncertainty about the current natural rate of unemployment and, hence, about the implied output gap.

Mark Gertler addressed two issues that he thought are being conflated. The first concerns what the output gap is, and the second is whether the current potential output is in part a response to demand contraction during the Great Recession. Regarding the first issue, Gertler agreed with Robert Hall and Valerie Ramey that the current output gap is low. Regarding the second issue, Gertler said that he was sympathetic to the view of the paper's authors, and he pointed out that the Great Recession looked similar to a financial crisis in an emerging market, with permanent deviations from output trends and permanent declines in productivity growth.

Glenn Rudebusch disagreed with those who considered the "laziness" of government economists to be an important factor responsible for the excess sensitivity of real-time estimates and potential output. He agreed with James Stock that measures of the unemployment gap are invariably quite persistent and smooth. If the output gap were to be set equal to the unemployment gap, then it would also be fairly smooth. However, there is notable noise or transitory variation in quarterly aggregate output. Therefore, transitory variation in potential output is a convenient mechanical offset that results in a smooth output gap. A more transparent accounting of measurement error and noise in measured output would help resolve this problem.

Wendy Edelberg, who was working at the CBO at the time of her remarks, discussed the CBO's experience of projecting potential output. She noted that the CBO has endeavored to not be overly influenced by recent movements in the weakness of total factor productivity when projecting potential output growth over the forthcoming decade. For its 10-year projections, the CBO has put more weight on the growth of total factor productivity before the last few years than its normal procedure would suggest. Edelberg discussed figure 1 of the paper, which shows the CBO's downward revisions to potential output since 2007. The line extended for potential output in the 2007 projection in an ocular regression looks as if it is a continuing trend from the data before 2007. But this is an illusion. The CBO's projection for 2007 predicted steeper potential output growth than since 2004. In 2007, the CBO was projecting a sizable pickup in hours growth that in retrospect seemed implausible and inconsistent with the demographic data. Therefore, for the first few years of downward revisions, the CBO had been incorporating the fact that its projections of hours growth were too strong, and they had little to do with the weaknesses in output growth that were being witnessed.

Edelberg also noted that the CBO projects potential output by building it up from data on total factor productivity, the labor market, and capital. One of the reasons it was being marked down was because the growth of capital services was weak. However, in relatively recent years, the persistent weakness of total factor productivity growth had been the major reason for downward revisions to potential output, with which the CBO had been grappling. The CBO projects that current potential total factor productivity growth is weak, consistent with recent incidences, but that potential total factor productivity will revert upward in the future, in line with long-term trends. Therefore, the CBO is projecting an improvement in potential output growth, which, Edelberg acknowledged, was based more on long-term trends than on developments in recent data. And the CBO was projecting potential output growth to improve from about 1.7 percent at the end of 2017 to almost 2 percent over the year. Discussing the paper, she noted that it would be hard to reconcile that the output gap is big and negative given all the other indicators in the economy. However, perhaps the real question is whether current estimates of the output gap are a good indicator of the behavior of potential output over the forthcoming 5 or 10 years. Edelberg referred to Glenn Rudebusch's hurricane analogy, wondering how much weight should be applied to temporary factors that hold down potential output for a short period, when considering output growth for longer periods.[7]

7. Glenn D. Rudebusch, Daniel Wilson, and Tim Mahedy, "The Puzzle of Weak First-Quarter GDP Growth," *Economic Letter 2015–16* (Federal Reserve Bank of San Francisco), 2015.

Alan Blinder agreed with James Stock and picked up on Wendy Edelberg's comments, noting that though there are uncertainties in everything, those in capital and labor are relatively easy to handle. Total factor productivity is the real challenge. There are three things to consider about total factor productivity growth. The first is that it is not constant and changes over time. The second is that these changes in total factor productivity growth are completely unpredictable. And the third is that it is very hard to recognize the changes when they happen. It took a long time to catch on to the productivity deceleration in the 1970s and to the productivity acceleration in the 1990s. It is therefore pretty much impossible to forecast productivity. Blinder observed that, for monetary policy purposes, potential GDP growth is forecasted for the next three years. He suggested that the authors consider whether anything beats the forecast that says that total factor productivity growth in the next three to five years will be similar to what it was in the last three to five years.

Kristin Forbes asked the authors if they had looked at past estimates of different agencies in real time to track the most accurate ones in hindsight, assuming that the authors' estimate of potential output is the best one. She wondered which estimates should be used to make a set of potential output estimates, if there is no time to replicate the authors' technology.

Olivier Coibion thanked the organizers; the commenters, Valerie Ramey, and Serena Ng; and the participants for their insightful comments. He stated that the paper does two things. First, it evaluates how existing real-time estimates of potential respond to shocks; and second, it asks if measures of potential can be created that do better along this metric. Coibion observed that almost all the comments focused on the second aspect, so he would do the same. Responding to a common comment about using information from inflation, he stated that though the authors did not cover it in their presentation, they discuss this extensively in the paper. One view of recent inflation dynamics, as suggested by Steve Braun, is that it reflects a broken or very flat Phillips curve, in which case inflation is uninformative about the output gap. In the paper, Coibion and his colleagues consider a second view, which is an expectation-augmented Phillips curve using household inflation expectations. As they have shown in previous work, this provides a stable Phillips curve with no missing disinflation. That Phillips curve can therefore successfully be used to infer an output gap. Because inflation remains well below inflation expectations, this Phillips curve implies an output gap in the same range as the other measures imply. Coibion stated that their results therefore were also consistent with inflation dynamics.

Coibion also agreed with the broader point about the usefulness of combining information from other sources. For example, consumption information can be used, and long-run restrictions can be combined with information about inflation to get more precise estimates of the output gap. Coibion noted that as Serena Ng emphasized, the lack of precision in the estimates is a major concern and, therefore, combining additional information would be useful. He stated that he and the other authors had attempted to understand the implications of off-the-shelf methods relative to the CBO's estimates. He concluded that they were surprised to find that, by and large, all the methods gave a similar answer about the evolution of the output gap relative to the start of the Great Recession.