

BPEA Conference Drafts, September 13-14, 2018

The Cyclical Sensitivity in Estimates of Potential Output

Olivier Coibion, University of Texas at Austin

Yuriy Gorodnichenko, University of California, Berkeley

Mauricio Ulate, University of California, Berkeley

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

THE CYCLICAL SENSITIVITY IN ESTIMATES OF POTENTIAL OUTPUT

Olivier Coibion
UT Austin
and NBER

Yuriy Gorodnichenko
UC Berkeley
and NBER

Mauricio Ulate
UC Berkeley

First Draft: May 11th, 2017
This Draft: August 13th, 2018

Abstract: The fact that declines in output since the Great Recession have 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 U.S. 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 or not. Some structural VAR methodologies can avoid these shortcomings. These approaches suggest a much more limited decline in potential output following the Great Recession.

JEL: E2, E3

Keywords: Potential output, output persistence.

An earlier version of the paper was prepared for the second annual conference of the National Bank of Ukraine. Gorodnichenko thanks the NSF for financial support. We are grateful to seminar participants at the College of William and Mary, UT Austin, the Central Bank of Chile, the Federal Reserve Bank of San Francisco and the NBER Summer Institute as well as Larry Ball, Olivier Blanchard, Serena Ng, Valerie Ramey, and Jim Stock for helpful comments and Alexander Kopoin (OECD) for sharing OECD data on potential output.

1 Introduction

The Great Recession was characterized not just by large declines in economic activity in most advanced economies, but also ones that have persisted for a decade with no sign of these affected economies “catching up” to previously expected trend levels. If anything, it is the trends that are now being revised down in light of the continuing inability of these economies to close the output gaps first generated in 2008. As illustrated in Figure 1 for the U.S., estimates of potential output have been systematically revised downward since the Great Recession, such that all of the current deviations of output from past estimates of potential are now being reinterpreted as permanent declines in the productive capacity of the economy. These large downward revisions imply that the output gap appears closed, and this absence of any remaining slack 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 U.S. as well as across a wide range of countries. Using a variety of institutional sources for estimates of potential gross domestic product (GDP), we find that real-time estimates of this variable respond to cyclical shocks that have no long-run effects on the economy and under-respond 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 (Fed) and from the Congressional Budget Office (CBO) for the U.S. as well as estimates from the International Monetary Fund (IMF) 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 professional forecasters (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, as well as measures of total factor productivity. 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, while 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 U.S. but across 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 counter-cyclical 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 does not differentiate between the underlying sources of changes in GDP, be they monetary, technology, etc. Thus, we can 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 Blanchard and Quah (1989) approach to identify supply and demand shocks can successfully generate real-time estimates of potential output that are consistent with theoretical predictions. Indeed, when the Blanchard and Quah (1989) 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, tax, 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 following the Great Recession. Our estimates imply that the gap between potential and actual output in the U.S. has increased by more than 6 log percentage points between 2007Q1 (when the gap was likely close to zero) and 2017Q1, leaving ample room for policymakers to close this gap through demand-side policies if they so chose to.

We find similar evidence of a large output gap using other methods to calculate measures of potential output, such as the ones proposed by Gali (1999) and Cochrane (1994) or one based on an estimated Phillips curve. All these methodologies give similar results, pointing to an increase in the gap of 5-10 percentage points between 2007Q1 and 2017Q1. 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.

This paper touches on several literatures. It is most directly tied to recent work since the Great Recession focusing on the possibility of hysteresis: cases where demand shocks lead to permanent effects on the level of economic activity. While there are many mechanisms that can generate such effects (e.g. less R&D during periods of low investment as in Anzoategui et al. (2016), Benigno and Fornaro (2017) and Moran and Queralto (2017)), empirical evidence on hysteresis remains scant, as emphasized in Blanchard (2017), with most estimates of monetary and government spending shocks being consistent with the null that these shocks have no permanent effects on GDP (see Nakamura and Steinsson 2017 and Ramey 2016 for reviews of the literature on monetary and government spending shocks). 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 expected to be long-lasting. Ball (2014) documents that for most advanced economies, much of the declines in output after the Great Recession have been matched with declines in estimates of potential output. Fatas and Summers (2016) 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 since estimates of potential GDP fail to adequately distinguish between the underlying sources of changes in GDP.

Our paper also relates to work on news shocks and beliefs about long-run productivity. A strand of literature studies how news about future productivity can have contemporaneous effects on economic activity long before the productivity changes actually occur (e.g. Beaudry and Portier 2006, Barsky and Sims 2011, 2012). In that spirit,

Blanchard et al. (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 in Blanchard et al. (2017).

A third literature that we build on focuses on the implications of real-time measurement of the output gap for monetary policy. Orphanides and 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 trend. 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 over-stimulated the economy in the 1970s because of mismeasurement of potential output, it is now at risk of under-stimulating the economy by underestimating the productive capacity of the economy.

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 economic shocks, consistent with models where agents are not fully informed about the state. By comparing the long-run response of GDP to 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 2 presents information about the estimates of potential output used in the paper. Section 3 presents our baseline estimates, using U.S. data, of how estimates of potential GDP respond to economic shocks. Section 4 extends these results to a broader range of countries. Section 5 presents some examples of how estimates of potential output can be improved. Section 6 concludes.

2. How Estimates of Potential Output Are Created (and Used)

A seminal description of potential output is in Okun's (1962) presidential address. While the notion of potential or natural levels of output had been discussed as far back as Wicksell (1898) or Keynes (1936), Okun (1962) provided a sharper definition than had been previously utilized as well as guidance about how to estimate potential output (Hauptmeier et al. 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, while potential GDP is related to the non-accelerating 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 of 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 in Mishkin (2007), there are three broad classes of methods to construct a measure of potential output: statistical, production function, and structural (DSGE-based). 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 autoregressive (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 (e.g., 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 (e.g., labor, capital, multifactor productivity) are plugged into the production function to obtain potential output. Since 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 (e.g., 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, since 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 (see Fernald et al. (2017) for an application of this approach to the dynamics of output in the post-Great Recession period).

Finally, structural approaches use dynamic stochastic general equilibrium (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 (e.g. Andres et al. 2005). This methodology is particularly model-dependent 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, there is no historical real-time data 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 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). Since our empirical strategy involves studying the response of real-time estimates of potential output to supply (long-lived) vs demand (transitory) shocks, we are adopting an interpretation of potential output which 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.

2.1. Congressional Budget office (CBO)

The CBO uses the production function approach for estimating potential output. As described in 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 approximately 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 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 nonfarm business sector 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 sectors of agriculture and forestry, and 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. Real-time CBO estimates of potential output are available since 1991 at the annual frequency and since 1999 at the semiannual frequency.

Estimates of potential output by the CBO play an important role in fiscal policy discussions in the U.S. 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, e.g. 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. Prior to 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 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 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 like what was observed after the Great Recession.

2.2. Federal Reserve

While preparing macroeconomic projections (historically known as Greenbook forecasts) for meetings of the Federal Open Market Committee (FOMC), the staff of the Federal Reserve Board constructs a measure of the output gap (that is, the difference between actual and potential output) to assist the FOMC’s members in their decision making. As pointed out by Edge and Rudd (2016), 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) 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*

¹ While 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 NIPA, CBO (2001, p.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.”

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, Fleischman and 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 their judgmental estimates of potential output. Their 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 9 macroeconomic series: real GDP, real gross domestic income, the unemployment rate, the labor-force participation rate, aggregate hours for the nonfarm business sector, 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 CPI 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 (e.g. 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 1969-1987 period are provided in Orphanides (2004). For this earlier period, the quality of the estimates is likely to be worse since the estimates sometimes had to be obtained from a variety of sources (e.g., the Council of Economic Advisors) 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 decision-making 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 (e.g. Orphanides 2001) as a reason why the Federal Reserve allowed inflation to rise during the 1970s, and Greenspan’s perception that potential output was growing unusually rapidly in the 1990s explains why monetary policymakers over this period were less concerned about inflation than they normally would have been given the low unemployment rates (Gorodnichenko and Shapiro 2007).

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.

2.3. International Monetary Fund (IMF)

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 in de Resende (2014), 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 approach to measuring potential output is methodologically similar to measures reported in Greenbooks, in the sense that they use a combination of different methods to compute potential output and aggregate them using a great deal of judgement. At the same time, the IMF staff often uses the Hodrick-Prescott filter and/or multivariate methods such as the ones described in Blagrave et al. (2015) to construct measures of potential output. The IMF provides potential output estimates for 27 countries (see Appendix Table 1 for the list of countries). Nowcasts and one-year-ahead forecasts are available for 2003-2016. Since 2009, the IMF also provides up to five-year-ahead forecasts for potential output.

Estimates of potential output can play an important role in IMF 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 structural imbalances are then the primary focus of policy reforms associated with countries receiving funds from the IMF during times of crisis.

2.4. Organization for Economic Cooperation and Development (OECD)

OECD 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 total factor productivity is assumed to converge to a certain degree among different countries in the medium run. As pointed out in OECD (2012): "*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, 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 Appendix Table 1). For many countries, nowcasts and one-year-ahead forecasts are available since 1989. Since 2005, the OECD also reports 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.

2.5. Consensus Economics

Consensus Economics, a survey of professional forecasters, does not provide estimates of potential output but they report forecasts for the growth rate of actual output from 1 to 10 years into the future. Since 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 Appendix Table 1). Given the wide range of forecasters included in 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.

2.6. 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). Consensus forecasts, in contrast, are at a different horizon and are for actual GDP 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 (Appendix Figure 1).

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 (Greenbook, CBO, IMF, OECD, 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 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 Greenbooks, as extended backward by Orphanides (2004). In addition, we plot several statistical approaches to estimating potential GDP, including a one-sided 5-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 Greenbook estimate 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 in Fernald (2012).

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

$$(\Delta \log Y_{t|t}^* - \Delta \log Y_{t|t-1}^*) = \alpha + \beta(\Delta \log Y_{t-1|t-1}^* - \Delta \log Y_{t-1|t-2}^*) + \text{error}_t \quad (1)$$

where $\Delta \log Y_{t|s}^*$ 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 Greenbook revisions of potential GDP. With CBO, the coefficient on lagged revisions is similar but not significantly different from zero. The results are different for international data, with coefficients on past OECD revisions being not different from zero while those on past IMF and Consensus Economics revisions exhibiting negative predictability.

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

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

3.1 Measures of economic shocks

There is a long literature on identifying shocks that potentially drive business-cycle and longer-term fluctuations, particularly for the U.S. (see Ramey 2016 for a survey). 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 certain abuse of terminology. All of 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.³

For supply shocks, we consider changes in total factor productivity (TFP), oil price shocks and tax shocks. The former are measured as in Fernald (2012), which 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 identified in Kilian (2009).⁴ For tax shocks, we use Romer and Romer (2010)’s 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.

³ Because the units of these shocks vary, we normalize all shocks to be mean zero and have unit variance.

⁴ 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.

We consider three identified demand shocks, all related to policy. The first are monetary policy shocks. For the U.S., 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 et al. (2017) and set values after the onset of the zero-bound equal to zero.⁵

The second type of demand shock we consider are the military spending news shocks of Ramey (2016). Using real-time measures of the expected future path of defense spending in the U.S., Ramey constructs a measure of the present discounted value of future defense expenditures 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 that spending following Auerbach and Gorodnichenko (2012a). 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 and 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 in 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 will 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 policy-makers explicitly rule out this channel and emphasize that, in their view, monetary policy has only transitory effects on GDP.⁶ Their estimates of potential GDP should therefore be invariant to monetary shocks.

3.2 Effects of Shocks on Actual Output and Estimates of Potential Output in the U.S.

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:

⁵ We also experimented with monetary policy shocks identified via recursive ordering of VAR residuals as in Bernanke and Blinder (1992) and we found similar results, as documented in Appendix Figure 5.

⁶ For example, in a speech on March 3, 2017, Janet Yellen stated “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.”

$$\Delta \log Y_t = \alpha + \sum_{k=0}^K \phi_k \epsilon_{t-k} + \text{error} \quad (2)$$

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 to Greenbook forecasts, semi-annually otherwise. Since 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.⁷ Impulse responses 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 $h = 2$, etc.⁸

For each impulse response, we include 66% 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 max horizon (8 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 Appendix Table 2, together with more information that we describe later.

We plot the responses of actual output to each type of shock in Figure 4. Panel A focuses on the three supply shocks. In response to a TFP shock, output immediately rises about 0.5% points and remains persistently higher by about that 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), we again find the expected responses of output. Contractionary monetary policy shocks push output down. The point estimates are much less precise than in 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

⁷ Since 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 in Pagan (1984).

⁸ For monetary policy shocks, we constrain $\phi_0 = 0$ to capture the minimum delay restriction.

are also generally delayed).⁹ Immediate spending shocks as in Auerbach and Gorodnichenko (2012a) 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.

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

$$\Delta \log Y_{t|t}^* = \alpha + \sum_{k=0}^K \phi_k \epsilon_{t-k} + \text{error} \quad (3)$$

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, we also include 66% 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 parenthesis). The p-values are also included in Panel A of 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 obtain 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 in Coibion and Gorodnichenko (2012, 2015b). However, the fact that estimates of potential GDP evolve very gradually after tax shocks (which occur only for large legislative tax changes that staff members at the Board would be well aware of) 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 tradeoff between stabilizing inflation (which rises after a

⁹ While 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.

negative oil supply shock thereby calling for higher interest rates) and closing the output gap (which is falling and calling for lower interest rates) are therefore perceiving an even starker tradeoff since the rise in the estimate of potential output makes the output gap seem even more negative.¹⁰ This result is not driven by the specific measure of oil supply shocks (we find a similar result with the Kilian (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 which they incorrectly attribute to stronger global demand for oil from e.g. 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 their effects on income are actually short-lived.

Our results are not limited to these specific examples of identified shocks. For example, we can identify supply and demand shocks jointly as in Blanchard and Quah (1989) by running a 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 (Appendix Figure 2). Importantly, because this identification explicitly imposes 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 in estimates of potential GDP. In short, across identification schemes, we find an over-response of real-time estimates of potential GDP to demand shocks and an under-response to supply shocks.

¹⁰ The pronounced decline in the *perceived* 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 contributed to the Great Inflation of the 1970s.

3.3. Robustness of Baseline Results for the U.S.

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 auto-distributed lag specifications to estimate IRFs as in Romer and Romer (2004), namely:

$$\Delta \log Y_t = \alpha + \sum_{j=1}^J \delta_j \Delta \log Y_{t-j} + \sum_{k=0}^K \phi_k \epsilon_{t-k} + \text{error} \quad (4)$$

using $J = 4$ and $K = 8$. Results are presented in 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 instead point toward 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 available across both the shocks and the Greenbook estimates of potential GDP (1969-2011). The results, presented in Figure 5, 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.¹¹ 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 Greenbooks also include forecasts and backcasts of potential GDP growth (2 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 to raise or lower the entire path of projected and past potential GDP growth in response to shocks (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, 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 5-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 section 2.1, 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

¹¹ When we apply the ADL specification to oil supply shocks over the whole sample, we find the same result.

therefore considerably less precisely estimated. However, the responses of the estimates of potential GDP are still quite precise. Qualitatively, we find that CBO 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.

The fact that CBO forecasts of *long-run* potential respond similarly to nowcasts of potential GDP addresses one possible issue raised in 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 in 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.

3. 4 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 (e.g., if they affect current capital stocks) but not long-run levels of potential output (as would be implied by e.g. monetary neutrality). This is unlikely to be the case, however, since the long-horizon CBO forecasts of potential GDP respond approximately 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. Since these types of methods fail to identify the different potential sources of 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 5-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 (and the p-values are included in Panel C of Appendix Table 2). When using the HP-filtered series, we can very closely replicate the response of estimated potential GDP after every shock.¹² 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.¹³ It is then quite natural for these series to respond to all shocks that affect GDP, even if these movements are transitory in nature. But this endogenous response to cyclical shocks should then 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 or not.

Another way to see how closely the HP-filter can mimic real-time estimates of potential GDP, as well as the potential dangers of doing so, is illustrated in Figure 8. In Panel A, we plot the time path of potential GDP that would have been estimated in real-time using the HP-filter during the Great Recession period. Specifically, for each quarter, we apply an HP-filter to 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 U.S., 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 Panel B of Figure 8, which replicates this exercise for the Great Depression using data from Ramey and 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 was 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 begin to climb back up. Unless one is prepared to

¹² 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 3-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 (since oil prices and production are endogenous). This could lead an HP-filter of real GDP to rise after an oil supply shock.

¹³ The best match of HP-filtered series comes with high values of λ (we use $\lambda=500,000$). 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 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 Appendix Figure 5.

entertain the idea that the Great Depression reflected negative supply shocks that were offset by positive supply shocks in the mid 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.¹⁴

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

The Great Recession was of course not limited to the U.S. and the persistence of output declines in most major advanced economies has also been associated with declines in their potential output, as documented in Ball (2014). Indeed, despite widespread lackluster growth by historical standards since the Great Recession, the World Bank recently estimated that advanced economies have an output gap of zero on average, indicating that the large downward revisions to potential output estimated by the CBO for the U.S. since 2007 extend to other advanced economies (World Bank 2018). To what extent do the cyclical patterns documented above in estimates of potential GDP generalize to other countries? In this section, we turn to cross-country estimates of potential GDP, both from international organizations as well as from professional forecasters. Using international data gives us many more observations and thus more statistical precision and power.

4.1 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.¹⁵

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:

$$\Delta \log Y_{j,t|t} = \alpha_j + \gamma_t + \sum_{k=0}^K \phi_k \epsilon_{j,t-k} + \text{error}_{t,j} \quad (5)$$

$$\Delta \log Y_{j,t|t}^* = \delta_j + \kappa_t + \sum_{k=0}^K \psi_k \epsilon_{j,t-k} + \text{error}_{t,j} \quad (6)$$

where j indicates the country and α_j, δ_j and γ_t, κ_t denote country and time fixed effects respectively. The time frequency is semi-annual, 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 done for the U.S. For productivity, we use innovations in labor productivity, after conditioning on

¹⁴ Papell and Prodan (2012) analyze large recessions in the U.S. and other countries using long samples. Consistent with our analysis of the Great Depression, they find that actual output eventually catches up with pre-recession 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.

¹⁵ We exclude Norway from our analysis because this country relies heavily on energy exports.

past changes in labor productivity as well as country and time fixed effects.¹⁶ For oil shocks, we continue to use the Kilian measure of oil supply shocks but interact it with a country-specific measure of oil sufficiency (from the International Energy Agency's (IEA 2017) World Energy Statistics and Balances, available via the OECD) to distinguish it from the time fixed effects.¹⁷ 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 1980Q1 until 2016Q4 or as available.¹⁸ Finally, fiscal shocks are differences between ex-post government spending and ex-ante forecasts of government spending from the OECD, following Auerbach and Gorodnichenko (2012b).

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 3 are included in the figure and summarized in 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 around one or one and a half years and are mostly gone after three years (we only show IRF's up to 4 semesters in the figure).

The effects of these shocks on potential GDP are consistent with those obtained for the U.S. 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 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 under-cyclical after productivity shocks and over-cyclical after demand shocks documented in the U.S.

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 U.S., 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

¹⁶ 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 terms of 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.

¹⁷ 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).

¹⁸ A group of countries is in the eurozone 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 eurozone. From this VAR, we obtain monetary policy shocks which we append to the shocks identified in the pre-euro period. We estimate VARs on the full sample.

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 as those found in the OECD sample. The responses of the IMF's estimated levels of potential GDP respond similarly as those from the OECD: they rise inertially after productivity shocks, and respond inertially as well 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. We also again include for comparison 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 U.S.: 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.

4.2 Private Long-Horizon Forecasts of GDP growth rate

In addition to forecasts from international policy organizations, we consider how private forecasters adjust their beliefs about the long-run GDP growth rate in response to shocks. While forecasts of potential GDP are not readily available, Consensus Economics provides forecasts of GDP at long-horizons on a semi-annual basis. To the extent that cyclical fluctuations in GDP should be complete within 5 or so years, 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 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 Appendix Table 1 for 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).

5 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 do 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 a basis for a satisfactory method to estimate potential output. Specifically, we first use the Blanchard and Quah (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.

5.1 Blanchard and Quah Approach to Estimating Potential Output

In this simple, proof-of-concept exercise, we follow Blanchard and Quah (1989, BQ henceforth) 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 demand 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 (e.g. Romer and Romer 2004, Ramey 2016).

Because BQ and others emphasize the importance of structural breaks, we use a rolling window of 120 quarters.¹⁹ When applying the BQ approach, we use real-time data to ensure that our results are not driven by information not available to the econometrician. In a particular quarter (say 1995Q1) we use the vintages of real output growth and unemployment rate that were available at that point in time (obtained from the FRB of Philadelphia's real time database for macroeconomists), estimate the SVAR with long run restriction using these series and then perform the historical decomposition on this 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

¹⁹ We would like the rolling window to be big for the long-run identifying restriction to work well, but 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.

to the last quarter in a rolling-window sample. The next quarter’s (1995Q2) historical decomposition data point is going to use vintages that were not available yet in 1995Q1, and the previous quarter’s (1994Q4) historical decomposition data point used vintages that contained less information and stopped in 1994Q4. 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 US potential GDP.

Heuristically, we decompose output growth rate as $\Delta \log Y_t = g + \Delta \log Y_t^p + \Delta \log Y_t^c$ where g is the long-run growth 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 forward, we construct forecasts for $\Delta \log Y_{t+h|t}^* = g + \Delta \log Y_{t+h|t}^*$ given the history of supply shocks up to period t . Then we cumulate $\Delta \log Y_{t+h|t}^*$ over $0, \dots, H$ 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. This assumption is consistent with the fact that the growth rate of output per capita in the U.S. has been stable at 2 percent per year over the last 150 years (Jones 2016). In our baseline analysis, we set g equal to the average growth rate of output in the post-WWII period.

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 regressions (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.

The fact that real-time estimates of potential output coming from the BQ do not suffer from the same issues identified as those found from official estimates of potential output is notable. 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.²⁰

²⁰ 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 (Appendix Figure 8). Despite a long delay, revised estimates of potential GDP from official agencies do not successfully distinguish between

The different estimates of potential output coming from the Blanchard and Quah (1989) methodology are also not innocuous in their policy implications relative to the available estimates of potential output. In Panel A of Figure 13, we plot the real-time revisions in potential output from the BQ methodology during the Great Recession. Like official estimates, we find that there are declines in potential output during the Great Recession that take 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 approximately 6 log percentage points more than actual output since 2007, a difference which could potentially be closed through the use of demand side policies.²¹

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. Since the BQ approach is estimated over a long period, more persistent demand shocks during the ZLB are likely to be in part attributed to “supply shocks” in the BQ decomposition. Some of the estimated decline in potential output since the Great Recession attributed to supply side factors is therefore likely to be transitory in nature, making the output gap even larger than our estimates suggest.

5.2 Alternative Estimates of Potential Output after the Great Recession

The BQ methodology points to a very different view of potential output since the Great Recession than that suggested by official and private organizations. In this section, we consider several alternative theory-based approaches to investigate the robustness of this finding.

One approach closely related to BQ is from Gali (1999). He proposes to identify 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 2-variable VAR as used in Gali (1999) on real-time data and define the real-time level of potential output as the level of output coming only from the identified technology shocks. As illustrated in Panel B of Figure 13, this approach points to even smaller revisions to the output gap over the course of the Great Recession, perhaps due to the narrower interpretation of the types of shocks that affect potential output

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.

²¹ One may be concerned that the full-sample estimate of g may be too optimistic when we compute trajectories of potential output. For example, our assumption about g is significantly higher than what the CBO assumed during the Great Recession, which explains why BQ forecasts of potential output for 2016 from 2007 or 2009 look high compared to CBO forecasts in Figure 1. In Panel A of Appendix Figure 9 we examine the sensitivity of projections to using the mean growth rate of GDP over the 1977-2007 period and we find similar results (output gap in 2016 in the 2017 vintage of BQ is close to 7 log percentage points). Relatedly, the dynamics of output growth may be influenced by low-frequency changes in the growth rate of population. Panel B of Appendix Figure 9 shows that normalizing output by population in the BQ approach yields a 4 log percentage point output gap in 2016 in the 2017 vintage of BQ.

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 2007.²²

Cochrane (1994) proposes an alternative approach to identifying permanent changes in GDP by exploiting the consumption/output ratio. Under the Friedman (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 Panel C of Figure 13. There is a large downward revision in the predicted path of potential output between 2007 and 2009 of about 5 percentage points, but subsequent revisions are very small. As with the Gali (1999) approach, the implied output gap in 2017 is therefore more than 10 log percentage points bigger than in 2007.

Importantly, the Cochrane approach is immune to concerns about hysteresis, since 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 slack.

Closer in spirit to Okun’s (1962) approach is to utilize information about potential output from the inflation rate. In New Keynesian models, nominal rigidities generate an expectations-augmented Phillips curve which 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 (2015a), we estimate an expectations-augmented Phillips curve during the pre-Great Recession period using inflation expectations from the Michigan Survey of Consumers. As shown in Coibion and Gorodnichenko (2015a), 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 implied to account for inflation dynamics during this time 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). Because there are few revisions to inflation data and we cannot forecast the

²² One could also follow King et al. (1991), Gonzalo and Ng (2001) and others to consider VARs that include more than two variables or use other permanent-transitory decompositions.

path of future potential GDP from the Phillips curve, we cannot replicate the previous results of plotting predicted paths of potential output from different periods. Instead, we plot a smoothed version of 2017 estimates of potential GDP over the period of the Great Recession in Panel D of Figure 13, along with the 2017 estimates from other approaches for comparison.²³ These estimates of potential GDP do 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 short, bringing additional information to bear on the identification of potential output, be it 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 approach. Each approach points to non-trivial revisions in potential output following 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.

This view is in sharp contrast to the conclusion one might reach from looking at recent U.S. unemployment rates, which fell below 4% in April 2018. Typical estimates of the NAIRU would point toward higher values (the 2018 CBO estimate is 4.6%), whereas our view that the output gap exceeds 5% would, using Okun's Law, imply a natural rate of unemployment of only around 1%, far too low to be realistic. Our interpretation is that some of the sharp declines in labor force participation rates since the Great Recession are cyclical and would be reversed if economic activity increased significantly. These unusual changes in labor force participation make the current unemployment rates unusually misleading indicators of actual labor market conditions. As illustrated in Figure 13, our interpretation is consistent with the behavior of recent inflation. In contrast, if the current unemployment rate was indeed significantly lower than the NAIRU, one would expect far more inflationary pressures and wage growth than have yet been seen.

5.3 Caveats

Okun (1962) 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 example, forecasts for actual output are routinely associated with wide confidence bands (e.g., standard errors for the Fed and private one-year-ahead forecasts are often greater than one percentage point). Since 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 BQ and similar methods in relatively short samples. Structural breaks and low-frequency variation in the data add another layer of complexity.

²³ 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.

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 per year, over ten years we can overstate the output gap by 1 percentage point.²⁴ As a result, because estimating potential output is inherently so challenging, one should interpret our estimates in this section as rather tentative.

6 Conclusion

Our results speak to two distinct but related questions. The first is how real time estimates of potential output respond to transitory vs. permanent economic shocks and therefore how we should interpret revisions in 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, be they 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 Blanchard and Quah (1989) that 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 IMF 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 these 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 the effects of tax changes on projected revenues is likely overstated. In this sense, our results are related to Blanchard and 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 public and private organizations to estimate potential GDP in real-time. Not only are there profound statistical and

²⁴ 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% per year between 2014 and 2020, well under the CBO's forecast of 2.2% a year, and far below the historical average of 3.1% (1947-2017 sample).

economic challenges involved, as described in section 5.3, 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 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 policy-making as well as proposing some 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 significant 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 estimated separately. Since 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 “[m]eanwhile, the measure of potential must be used with care.”

References

Andres, Javier, David Lopez-Salido, and Ed Nelson, 2005. “Sticky-Price Models and the Natural Rate Hypothesis,” *Journal of Monetary Economics* 52(5): 1025-1053.

Anzoategui, Diego, Diego Comin, Mark Gertler, and Joseba Martinez, 2016. “Endogenous Technology Adoption and R&D as Sources of Business Cycle Persistence,” NBER Working Papers, 22005.

Auerbach, Alan, and Yuriy Gorodnichenko, 2012a. “Measuring the Output Responses to Fiscal Policy,” *American Economic Journal – Economic Policy* 4(2): 1–27.

Auerbach, Alan, and Yuriy Gorodnichenko, 2012b. “Fiscal Multipliers in Recession and Expansion,” in *Fiscal Policy after the Financial Crisis*, Alberto Alesina and Francesco Giavazzi, eds., University of Chicago Press.

Ball, Laurence, 2014. “Long-Term Damage from the Great Recession in OECD Countries,” Working Paper, John Hopkins University, May 2014.

Barsky, Robert, and Eric Sims, 2011. “News Shocks and Business Cycles,” *Journal of Monetary Economics* 58(3): 273-289.

Baumeister, Christiane, and Gert Peersman, 2013. “Time-Varying Effects of Oil Supply Shocks on the US Economy,” *American Economic Journal: Macroeconomics* 5(4): 1-28.

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, November 2015.

Beaudry, Paul, and Franck Portier, 2006. “Stock Prices, News and Economic Fluctuations,” *American Economic Review* 96(4): 1293-1307.

Benigno, Gianluca, and Luca Fornaro, 2017. “Stagnation traps,” ECB Working Paper 2038.

Bernanke, Ben S., and Alan S. Blinder, 1992. "The Federal Funds Rate and the Channels of Monetary Transmission," *American Economic Review* 82(4): 901-921.

Blagrove, Patrick, et al., 2015. "A Simple Multivariate Filter for Estimating Potential Output." *IMF Working Paper*, 2015-79.

Blanchard, Olivier, 2017. "Should One Reject the Natural Rate Hypothesis?" Manuscript.

Blanchard, Olivier, and Daniel Leigh, 2013. "Growth Forecast Errors and Fiscal Multipliers," *American Economic Review* 103(3): 117-120.

Blanchard, Olivier, and Danny Quah, 1989. "The Dynamic Effects of Demand and Supply Disturbances," *American Economic Review* 79(4): 655-673.

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.

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.

Clark, Peter, 1979. "Potential GNP in the United States 1948-80," *Review of Income and Wealth*, 25: 141-165.

Clark, Peter, 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(1): 241-265.

Coibion, Olivier, and Yuriy Gorodnichenko, 2012. "What Can Survey Forecasts Tell Us about Information Rigidities?" *Journal of Political Economy* 120(1): 116-159.

Coibion, Olivier, and Yuriy Gorodnichenko, 2015a. "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation," *American Economic Journal: Macroeconomics* 7(1): 197-232.

Coibion, Olivier, and Yuriy Gorodnichenko, 2015b. "Information Rrigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review* 105(8): 2644-2678.

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.

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

Congressional Budget Office, 2014. Revisions to CBO's Projection of Potential Output Since 2007. February 2014.

de Resende, Carlos, 2014. "An Assessment of IMF Medium-Term Forecasts of GDP Growth." *Background Paper*, Independent Evaluation Office of the International Monetary Fund. February 12, 2014.

Edge, Rochelle, and Jeremy Rudd, 2016. "Real-time Properties of the Federal Reserve's Output Gap." *Review of Economics and Statistics* 98(4): 785-791.

Fatas, Antonio, and Summers, Lawrence, 2016. "The Permanent Effects of Fiscal Consolidations." Manuscript.

Fernald, John G., 2007. "Trend breaks, long-run restrictions, and contractionary technology improvements," *Journal of Monetary Economics*, Elsevier, vol. 54(8), pages 2467-2485, November.

Fernald, John G., 2012. "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity." *FRBSF Working Paper*, 2012-19 (updated April 2014).

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 2017.

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

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

Gali, Jordi, 1999. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review* 89(1): 249-271.

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(10): 1527-1546.

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.

Gordon, Robert, and Robert Krenn, 2010. "The End of the Great Depression 1939-41: Policy Contributions and Fiscal Multipliers," *NBER Working Paper*, 16380.

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(4), 1132-1162.

International Energy Agency (IEA), 2017. "World Indicators," IEA World Energy Statistics and Balances (database). DOI: <http://dx.doi.org/10.1787/data-00514-en>. (Accessed on 12 May 2017).

Jones, Charles I., 2016. "The Facts of Economic Growth," in John B. Taylor and Harald Uhlig, eds., *Handbook of Macroeconomics*, Elsevier, Volume 2, pp. 3-69.

Keynes, John M., 1936. *A General Theory of Employment, Interest and Money*. 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(2): 216-240.

Kilian, Lutz, 2009. "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market." *American Economic Review* 99(3): 1053-1069.

King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson, 1991. "Stochastic Trends and Economic Fluctuations," *American Economic Review* 81(4): 819-840.

Kurmann, Andre and Eric Sims, 2017. "Revisions in Utilization-Adjusted TFP and Robust Identification of News Shocks," manuscript.

Kuttner, Kenneth, 1994. "Estimating Potential Output as a Latent Variable," *Journal of Business and Economic Statistics* 12(3): 361-368.

Mishkin, Frederick, 2007. *Estimating Potential Output. Speech at the Conference on Price Measurement for Monetary Policy*, Federal Reserve Bank of Dallas. Dallas, Texas. May 24, 2007.

Moran, Patrick and Albert Queralto, 2017. "Innovation, Productivity and Monetary Policy." *Journal of Monetary Economics*, forthcoming.

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

Organization for Economic Cooperation and Development (OECD), 2012. *OECD Economic Outlook*. Chapter 4, Medium and long-term scenarios for global growth and imbalances. Volume 2012/1.

Orphanides, Athanasios and Simon van Norden, 2002. "The Unreliability of Output-Gap Estimates in Real Time," *Review of Economics and Statistics* 84(4): 569-583.

Orphanides, Athanasios, 2003. "The quest for prosperity without inflation." *Journal of Monetary Economics* 50: 633-663.

Orphanides, Athanasios, 2004. "Monetary Policy Rules, Macroeconomic Stability, and Inflation: A View from the Trenches." *Journal of Money, Credit and Banking* 36(2): 151-175.

Orphanides, Athanasios. 2001. "Monetary Policy Rules Based on Real-Time Data." *American Economic Review* 91(4): 964-985.

Pagan, Adrian, 1984. "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review* 25(1): 221-247.

Papell, David, and Ruxandra Prodan, 2012. "The Statistical Behavior of GDP after Financial Crises and Severe Recessions," *B.E. Journal of Macroeconomics* 12(3): 1-31.

Ramey, Valerie A., and Sarah Zubairy, 2018. "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data," *Journal of Political Economy* 126(2): 850-901.

Ramey, Valerie, 2016. "Macroeconomic Shocks and Their Propagation," in *Handbook of Macroeconomics, Volume 2*, John Taylor and Harald Uhlig, eds., Elsevier Science.

Romer, Christina, and David Romer, 2004. "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review* 94(4): 1055-1084.

Romer, Christina, and David Romer, 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review* 100(3): 763-801.

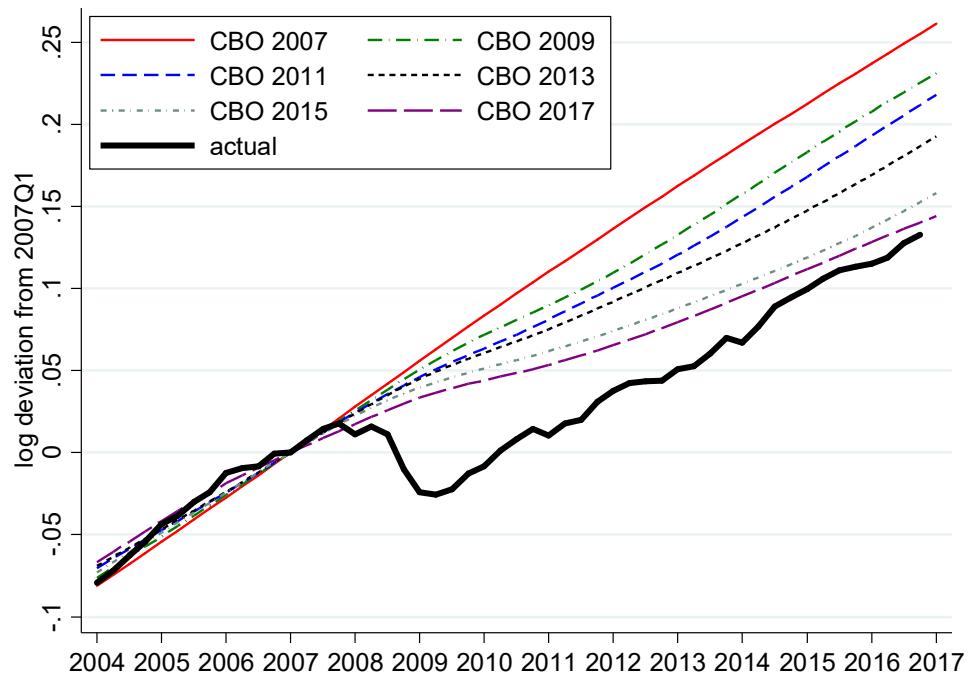
Staiger, Douglas O., James H. Stock, and Mark W. Watson, 1997. "How Precise Are Estimates of the Natural Rate of Unemployment?," in: *Reducing Inflation: Motivation and Strategy*, pages 195-246, National Bureau of Economic Research, Inc.

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

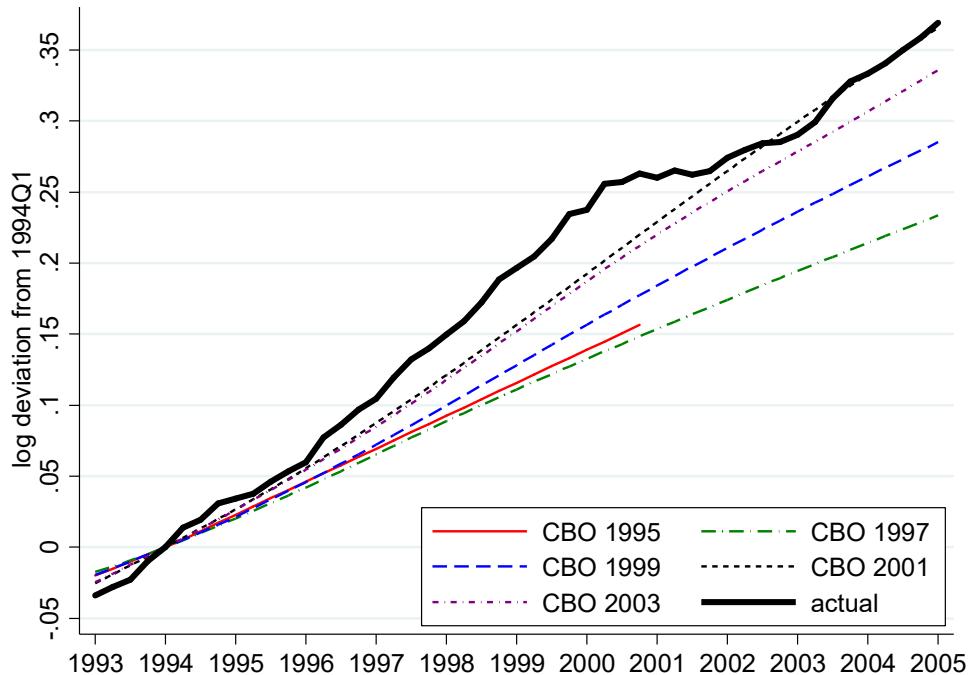
World Bank, 2018. "Global Economic Prospects: Broad Based Upturn but for How Long?" January 2018.

Figure 1: Historical Revisions in CBO Estimates of U.S. Potential Output.

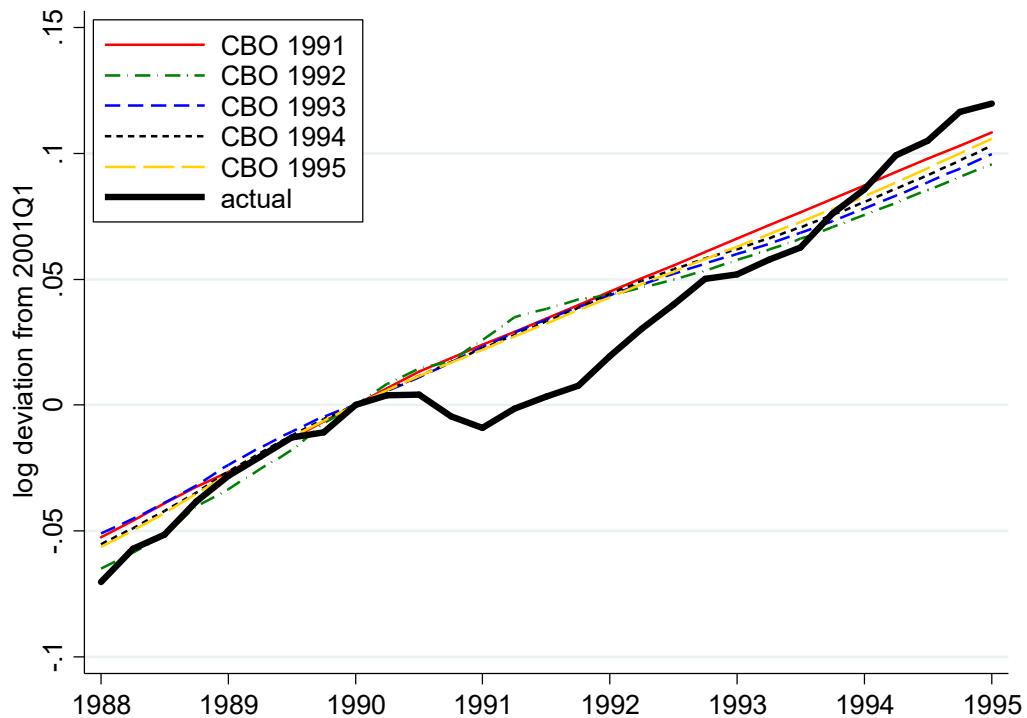
Panel A: The Great Recession



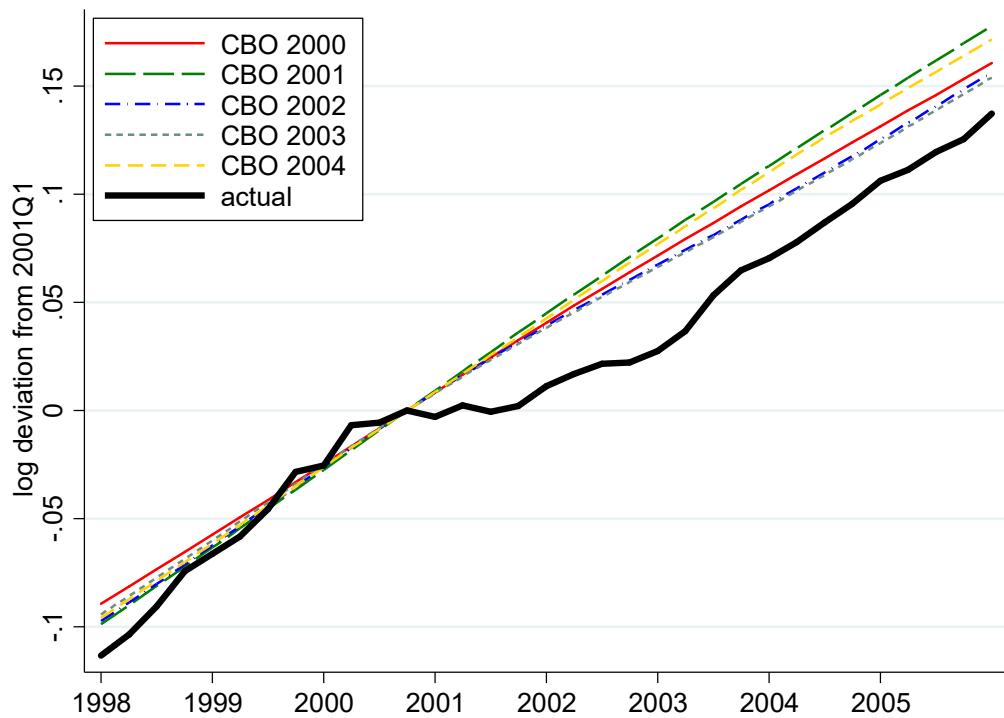
Panel B: The 1990s Productivity Boom



Panel C: The 1990 Recession

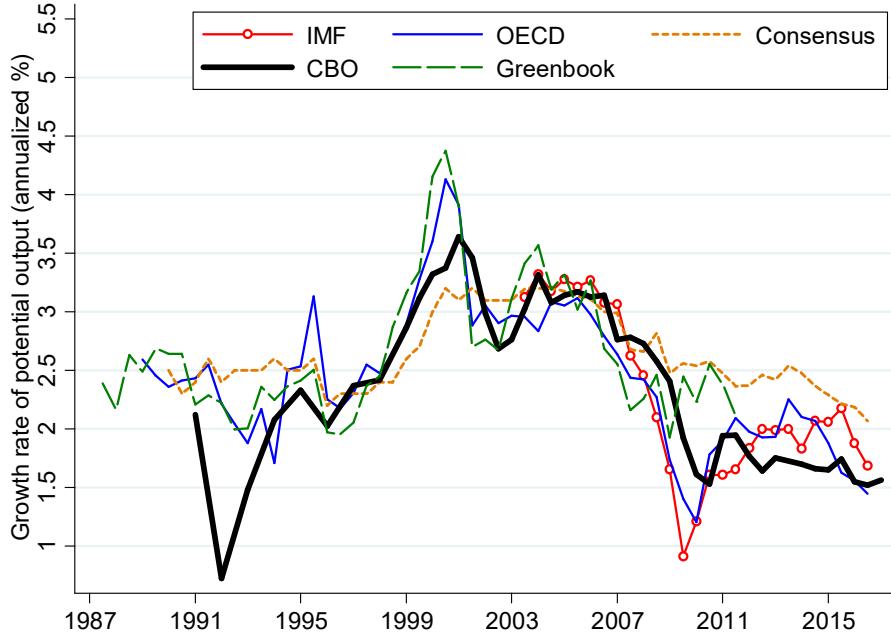


Panel D: The 2001 Recession



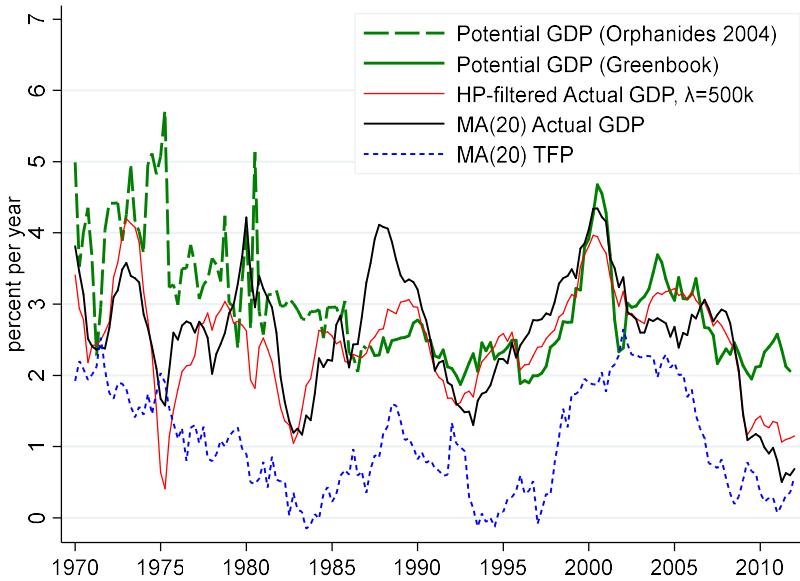
Notes: The figure plots estimates of U.S. potential output from the Congressional Budget Office made at different time periods (beginning of the corresponding year). The solid black line represents real GDP in the U.S.

Figure 2: Estimates of potential output growth rate and forecasted long-term growth for actual output, USA.



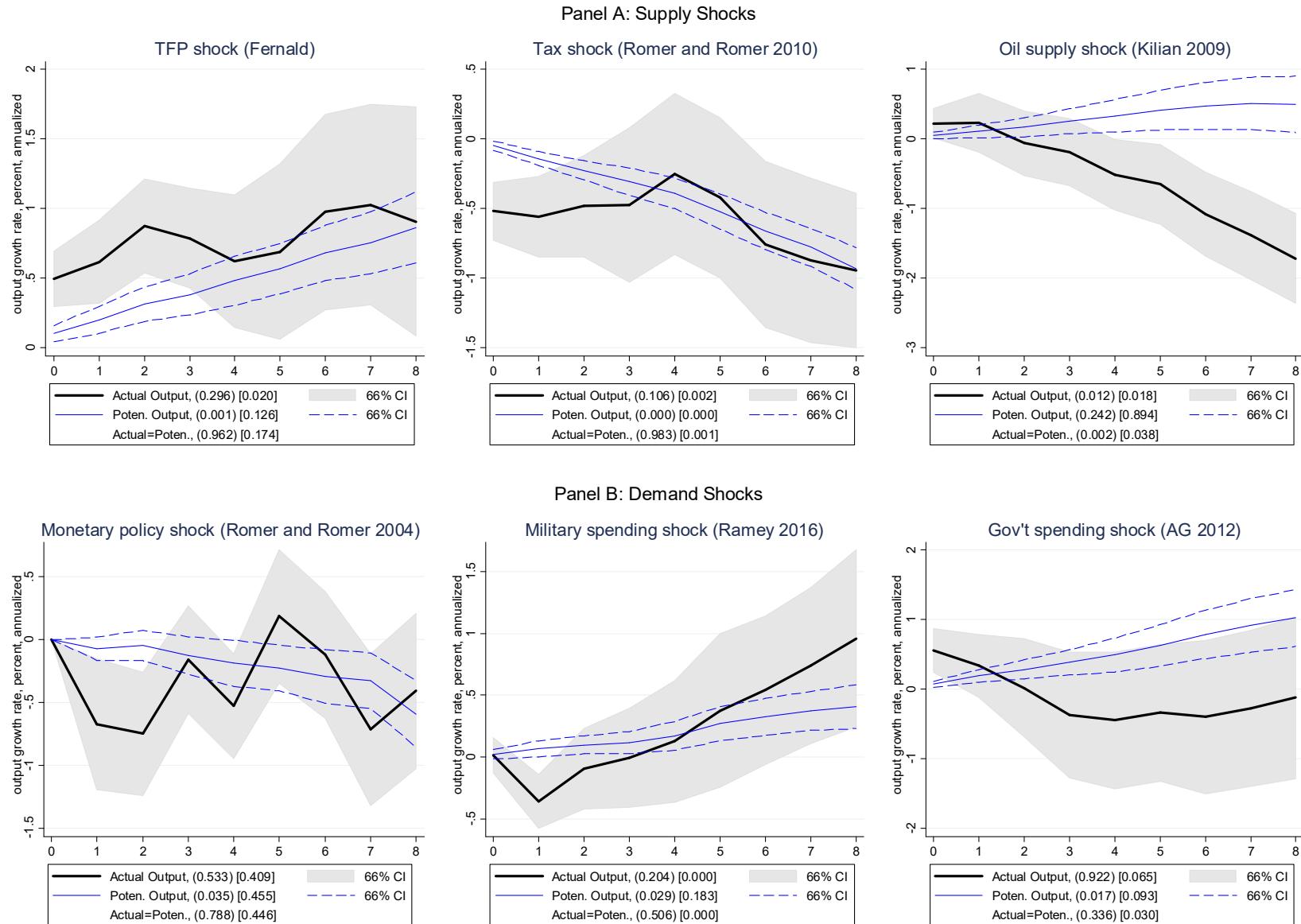
Notes: All series in the figure are real time data at the semi-annual frequency. The potential output for IMF, OECD, and CBO is reported for the current calendar year. Potential output for Greenbooks is the semiannual average of quarterly growth rates of potential output for the quarters in a given semester.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

Figure 3: Real-time estimates of potential output growth rate and trends in actual output growth rate, USA.



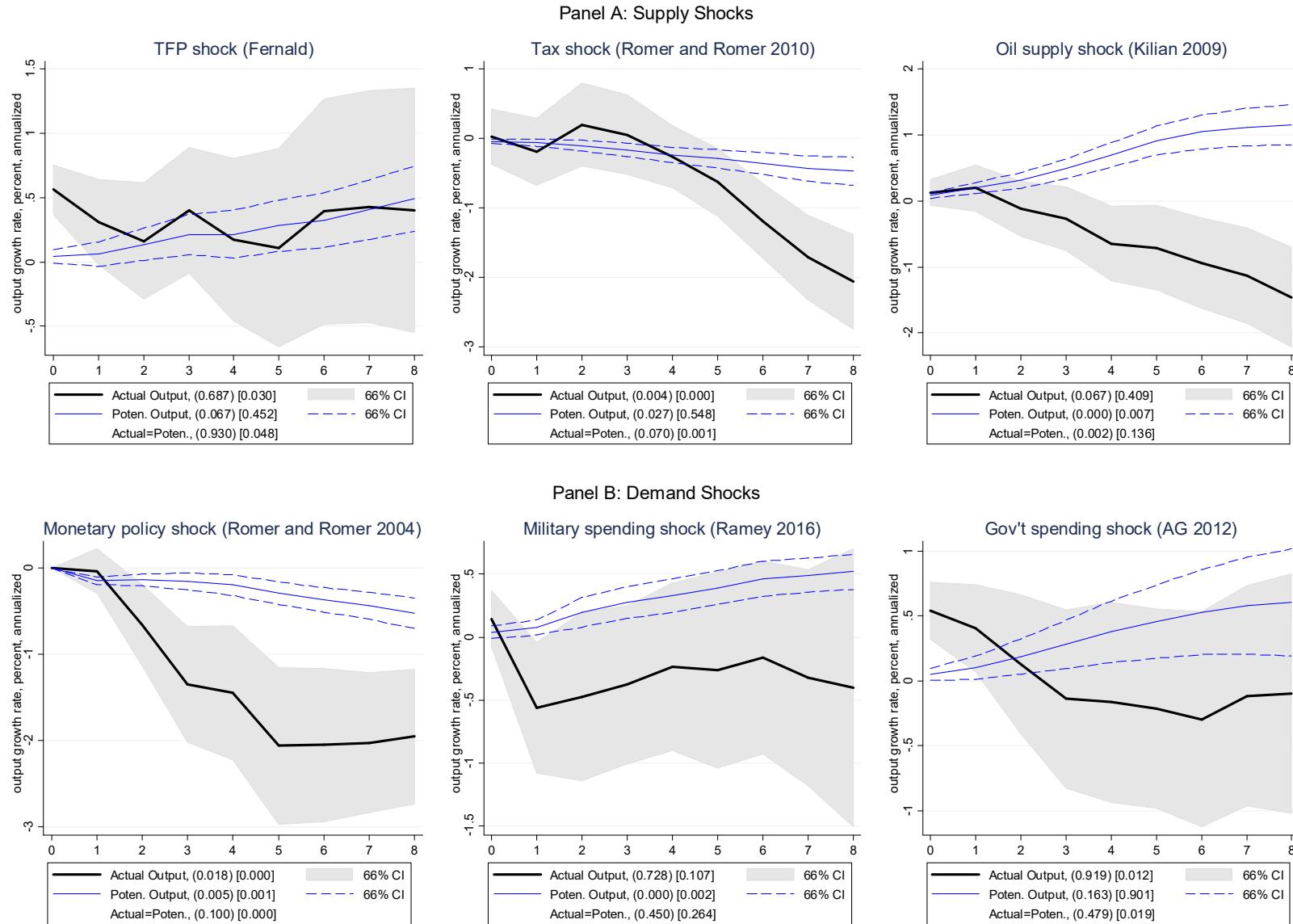
Notes: All series are real-time at the quarterly frequency. Potential output for the pre-1987 period is from Orphanides (2004). Potential output for 1987-2011 is from the Federal Reserve Bank of Philadelphia. Potential output is measured as the growth rate of potential output between a given quarter and the next 3 quarters. *HP-filtered actual output* 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) TFP* for a given quarter is calculated as the 20-quarter moving average running on the current quarter and the preceding 19 quarters.

Figure 4: Responses of Output and Greenbook Estimates of Potential Output in U.S. to Shocks.



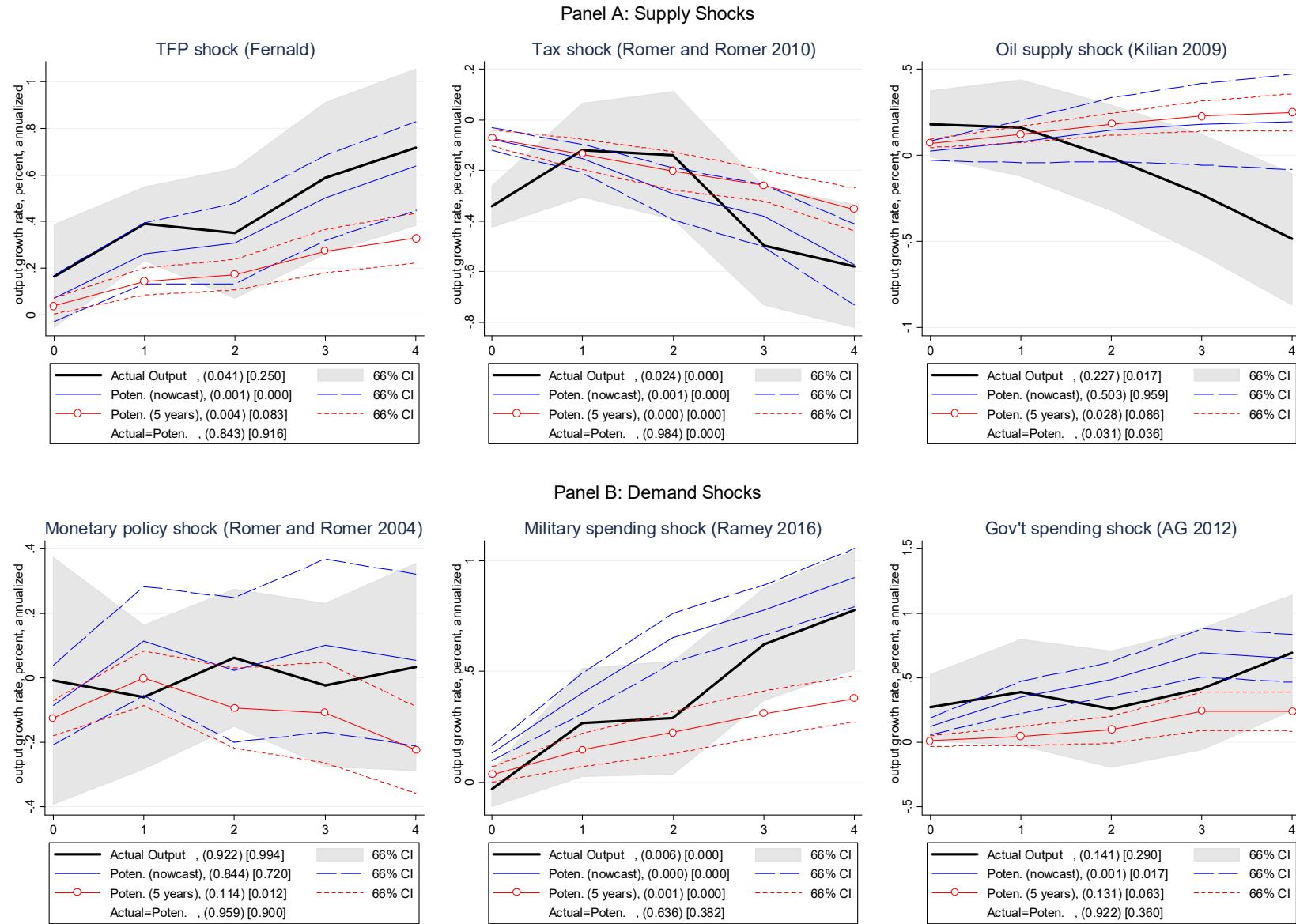
Notes: The figure reports impulse response functions (IRFs) estimated using equations (2) and (3). The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) available at the Federal Reserve Bank of Philadelphia. In parentheses we report the p-value for a test of whether the response of actual (potential) output is different from zero at the max horizon (8 quarters), while in square brackets we show the p-value 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 reports p-values for a test of equality of responses of actual and potential output at the max horizon (parentheses) and a test of equality of the paths of the responses for actual and potential output are equal across horizons.

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



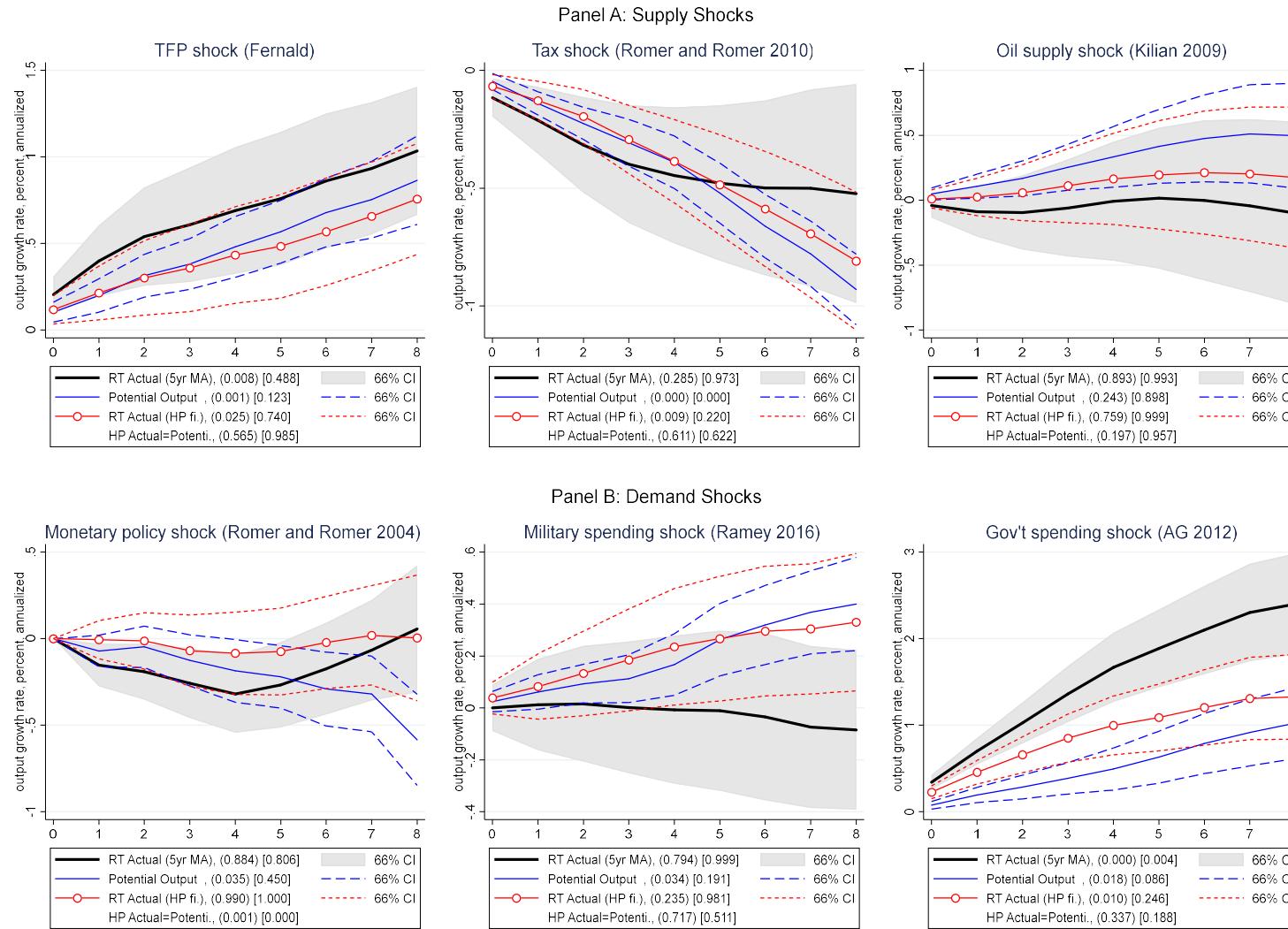
Notes: The figure reports impulse response functions (IRFs) estimated using equation (2) and (3). The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) using output gap data starting in 1970. In parentheses we report the p-value for a test of whether the response of actual (potential) output is different from zero at the max horizon (8 quarters), while in square brackets we show the p-value 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 reports p-values for a test of equality of responses of actual and potential output at the max horizon (parentheses) and a test of equality of the paths of the responses for actual and potential output are equal across horizons.

Figure 6: Responses of Output and CBO Estimates of Potential Output in U.S. to Shocks.



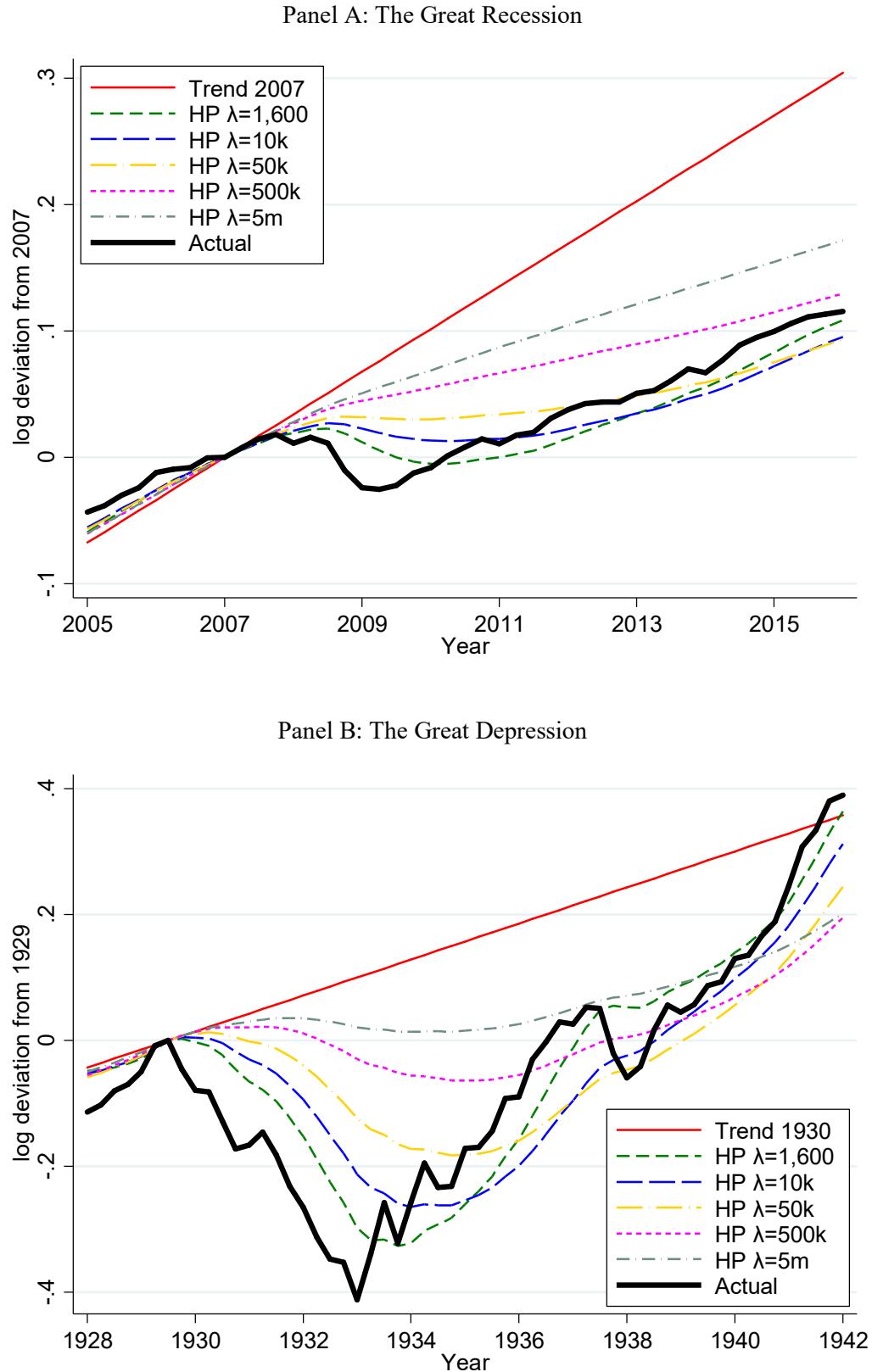
Notes: The figure reports impulse response functions (IRFs) estimated using equations (2) and (3). The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) available from the Congressional Budget Office. In parentheses we report the p-value for a test of whether the response of actual (potential) output is different from zero at the max horizon (8 quarters), while in square brackets we show the p-value 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 reports p-values for a test of equality of responses of actual and potential output (nowcast) at the max horizon (parentheses) and for a test of equality of the paths of the responses for actual and potential (nowcast) output are equal across horizons.

Figure 7: Responses of Greenbook Estimates of Potential Output and HP-filtered Output in U.S. to Shocks.



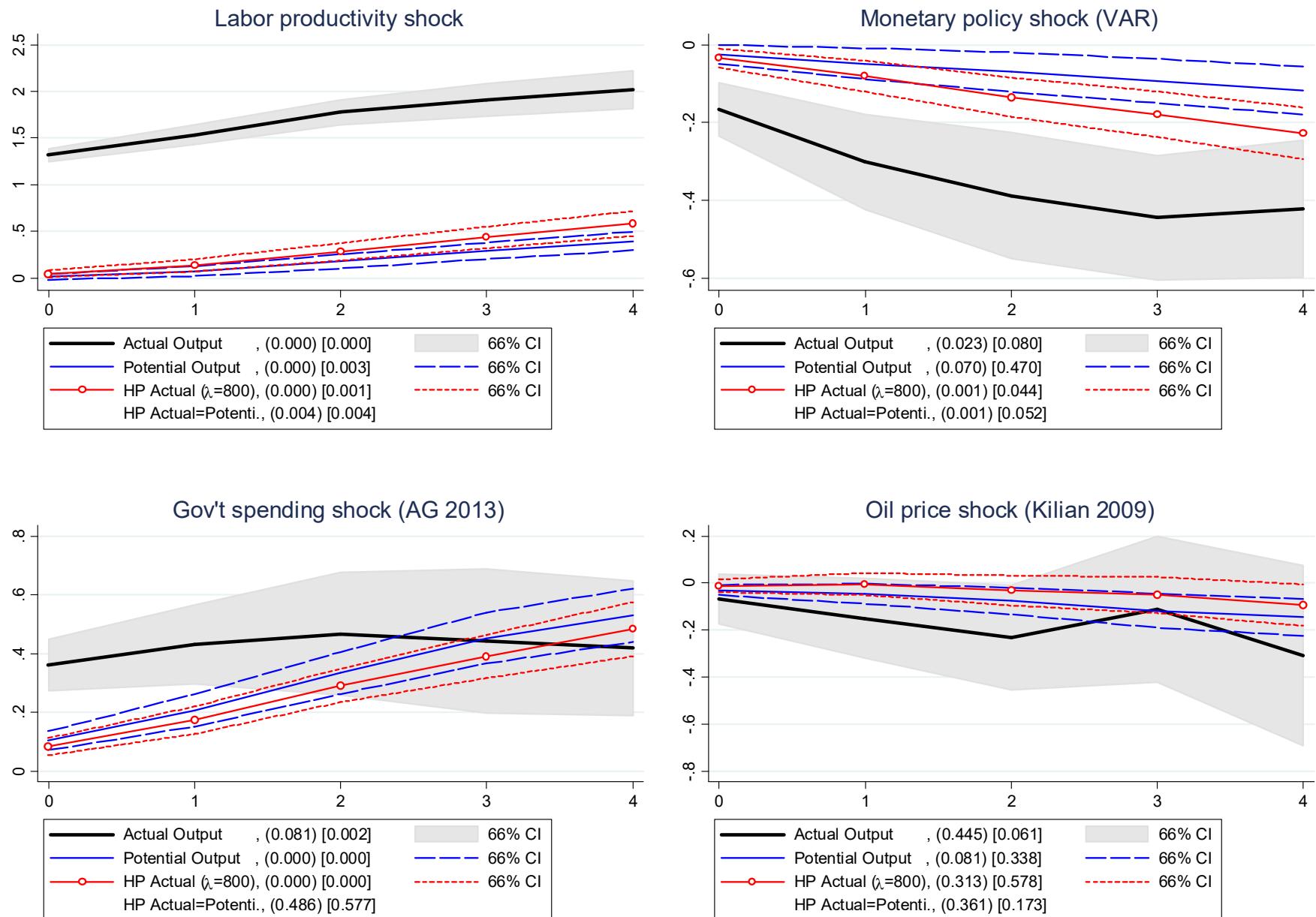
Notes: The figure reports impulse response functions (IRFs) estimated using equations (2) and (3). The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) available at the Federal Reserve Bank of Philadelphia. HP-filtered actual 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. 5-year moving average (MA) actual output for a given quarter is calculated as the 20-quarter moving average 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-value for a test of whether the response of actual (potential) output is different from zero at the max horizon (8 quarters), while in square brackets we show the p-value 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 reports p-values for a test of equality of responses of actual and potential output at the max horizon (parentheses) and for a test of equality of the paths of the responses for actual and potential output are equal across horizons.

Figure 8. Real-Time Estimates of Potential GDP from an HP-filter during Great Recession and Great Depression



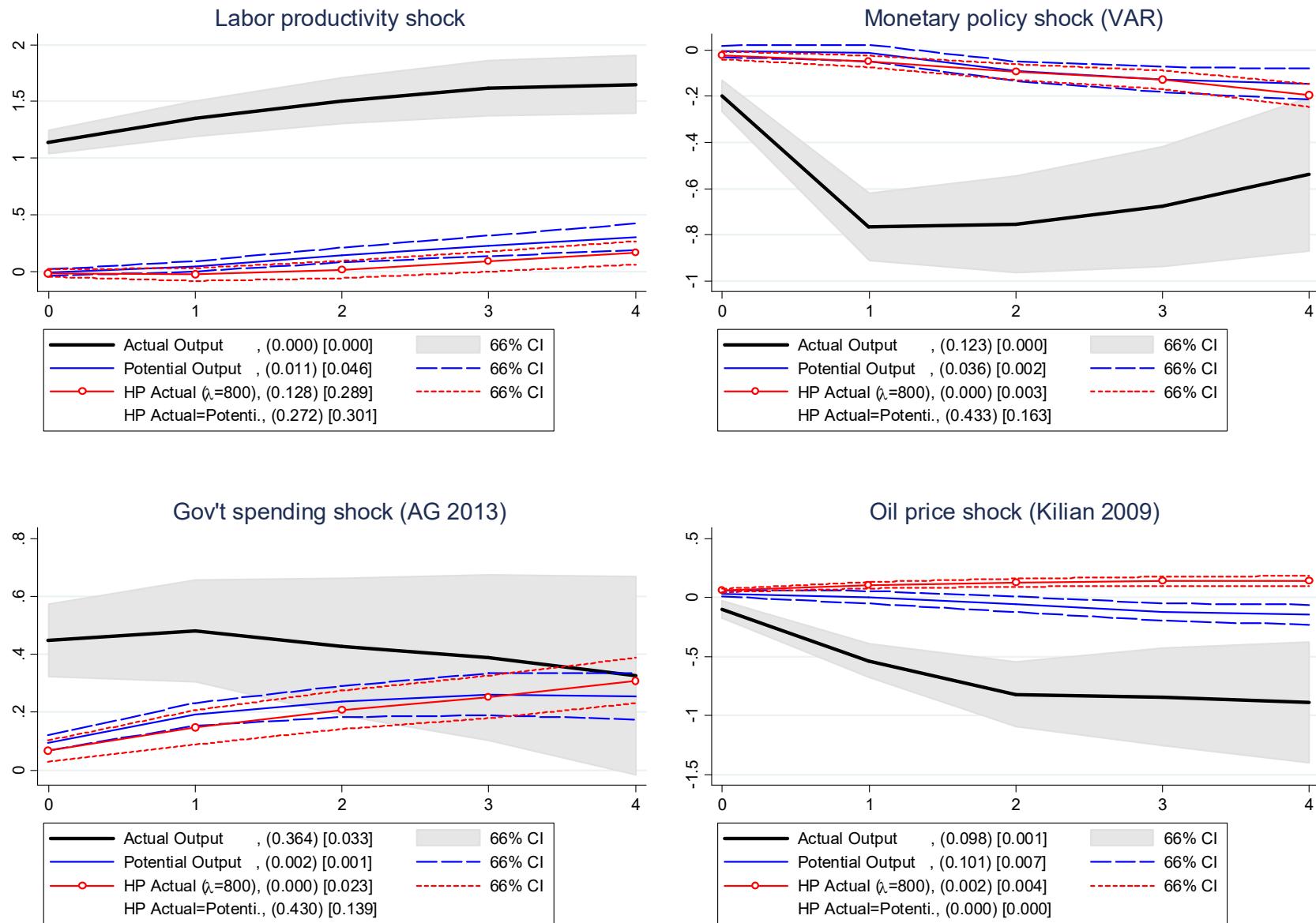
Notes: The figure report estimates of trend (potential output) generated by the one-sided Hodrick-Prescott filter for various values of the smoothing parameter λ . The filter is recursively applied to the final vintage of the data. For example, an estimate for 2008Q1 uses data only up to 2008Q1, an estimate for 2008Q2 uses data only up to 2008Q2, etc. Data in Panel B are from Ramey and Zubairy (2018).

Figure 9. Response of the growth rate for actual output and OECD's measure of potential output (nowcast).



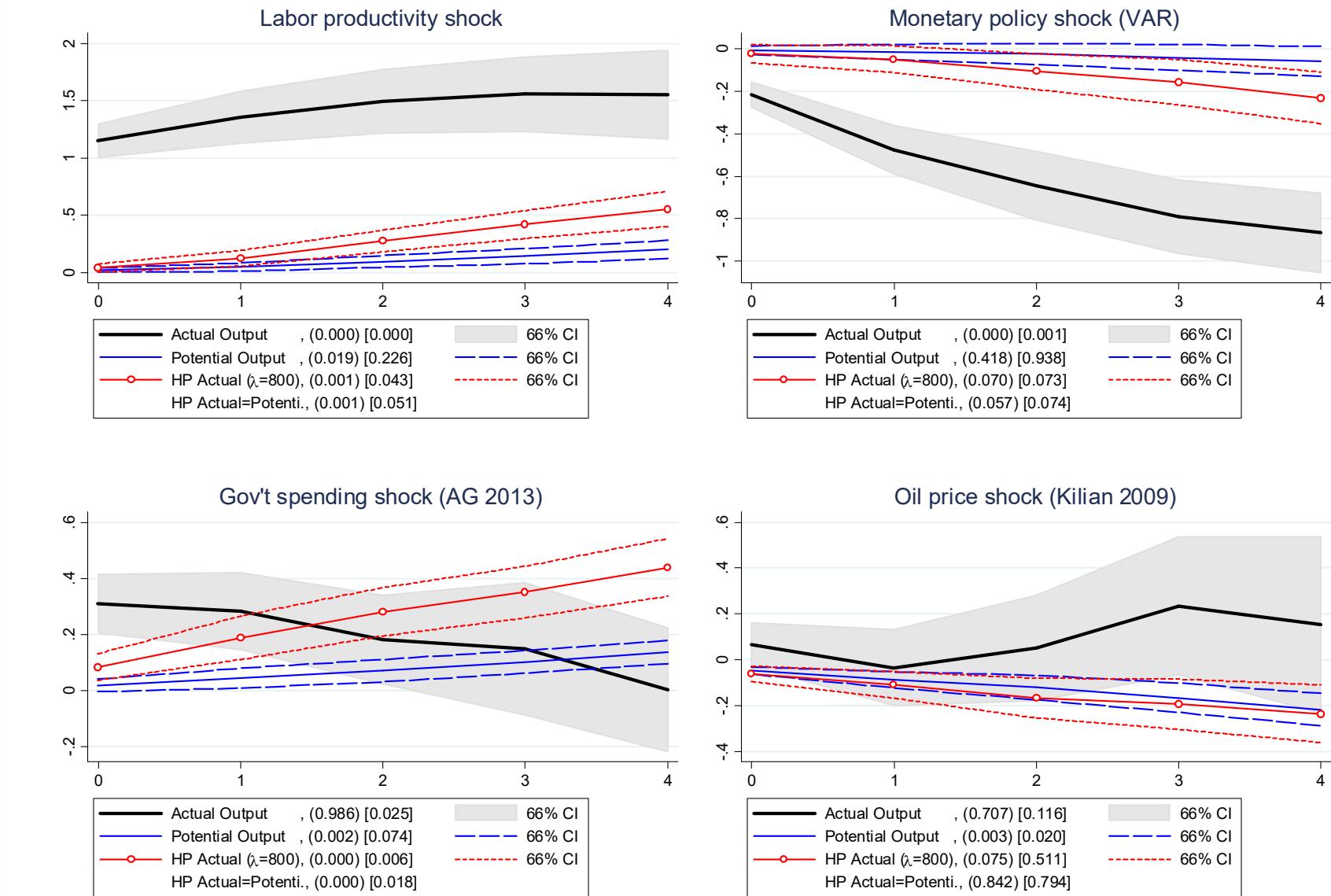
Notes: The figure shows impulse response functions (IRFs) for growth rates of actual and potential output (nowcast). IRFs are estimated using equations (5) and (6). The horizontal axis measures time in semesters (6 months). The vertical axis measures growth rate of output per year. In parentheses we report the p-value for a test of whether the response of actual (potential) output is different from zero at the max horizon (8 quarters), while in square brackets we show the p-value 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 reports p-values for a test of equality of responses of actual and potential output at the max horizon (parentheses) and for a test of equality of the paths of the responses for actual and potential output are equal across horizons.

Figure 10. Response of the growth rate for actual output and IMF's measure of potential output (nowcast).



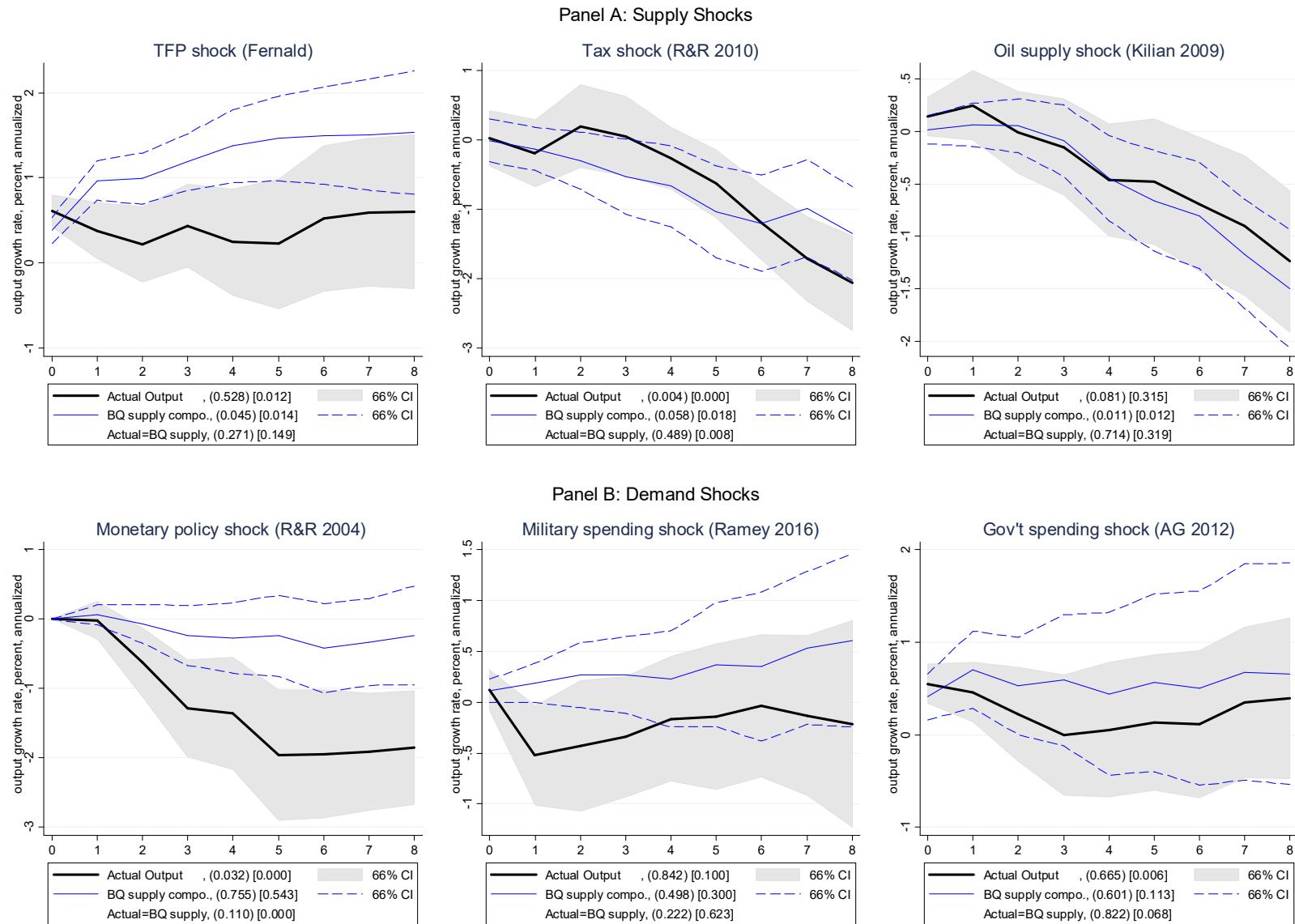
Notes: The figure shows impulse response functions (IRFs) for growth rates of actual and potential output (nowcast). IRFs are estimated using equations (5) and (6). The horizontal axis measures time in semesters (6 months). The vertical axis measures growth rate of output per year. In parentheses we report the p-value for a test of whether the response of actual (potential) output is different from zero at the max horizon (8 quarters), while in square brackets we show the p-value 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 reports p-values for a test of equality of responses of actual and potential output at the max horizon (parentheses) and for a test of equality of the paths of the responses for actual and potential output are equal across horizons.

Figure 11. Response of the growth rate for actual output and Consensus Economics' 6-10-year ahead forecast for actual output.



Notes: The figure shows impulse response functions (IRFs) for growth rates of actual output and 6-10-year ahead forecast for actual output growth rate (Consensus Economics). IRFs are estimated using equation (5) and (6). The horizontal axis measures time in semesters (6 months). The vertical axis measures growth rate of output per year. In parentheses we report the p-value for a test of whether the IRF of actual (potential) output is different from zero at the max horizon (8 quarters), while in square brackets we show the p-value for a test of whether the path of the IRF of actual (potential) output is different from zero across all horizons of the IRF. The last row of the legend reports p-values for a test of equality of responses of actual and potential output at the max horizon (parentheses) and for a test of equality of the paths of the responses for actual and potential output are equal across horizons.

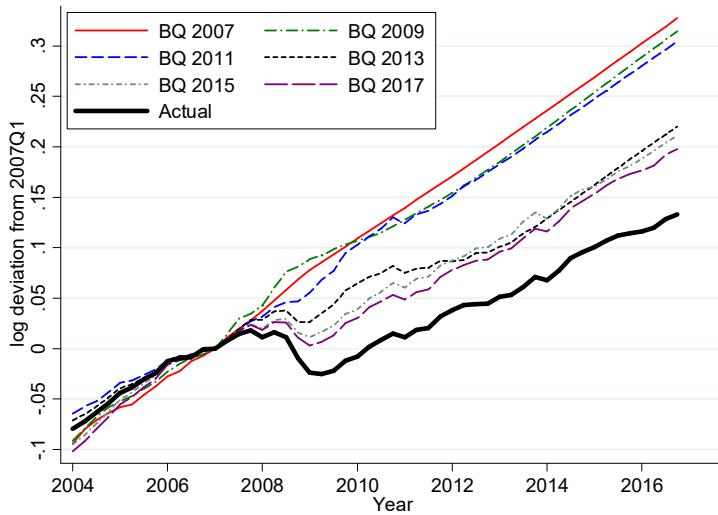
Figure 12: Response of the growth rate for actual output and SVAR identified historical supply component of actual output.



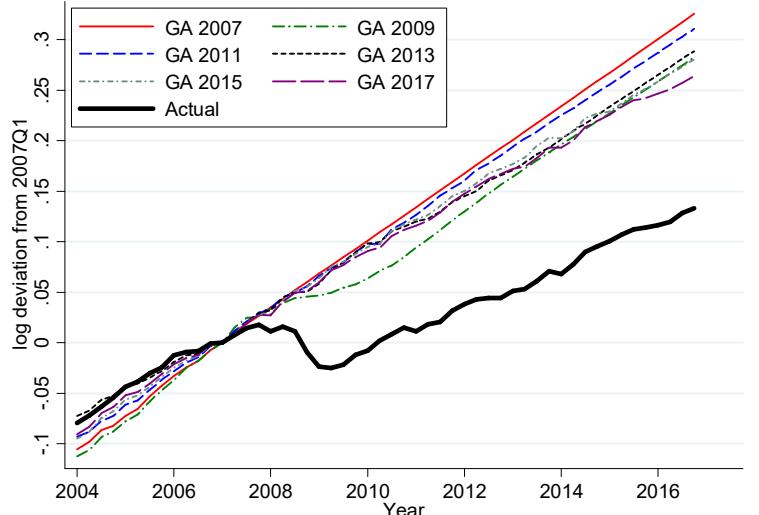
Notes: The figure reports impulse response functions (IRFs) estimated using equation (2) and (3). The “BQ Supply compo.” is the historical contribution of supply-side shocks (identified as in Blanchard and Quah 1989) to output growth rate. The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) using output gap data starting in 1970. In parentheses we report the p-value for a test of whether the response of actual (potential) output is different from zero at the max horizon (8 quarters), while in square brackets we show the p-value 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 reports p-values for a test of equality of responses of actual and potential output at the max horizon (parentheses) and for a test of equality of the paths of the responses for actual and potential output are equal across horizons.

Figure 13: Alternative Approaches to Estimating Potential GDP in Real-Time during the Great Recession

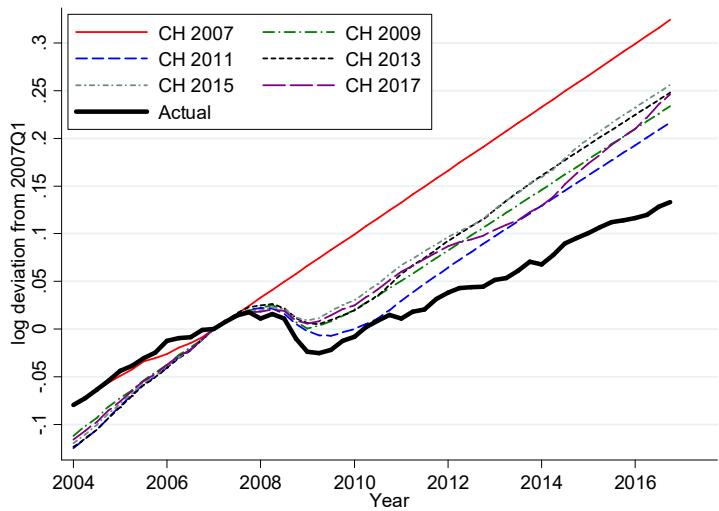
Panel A: Revisions using Blanchard and Quah (1989)



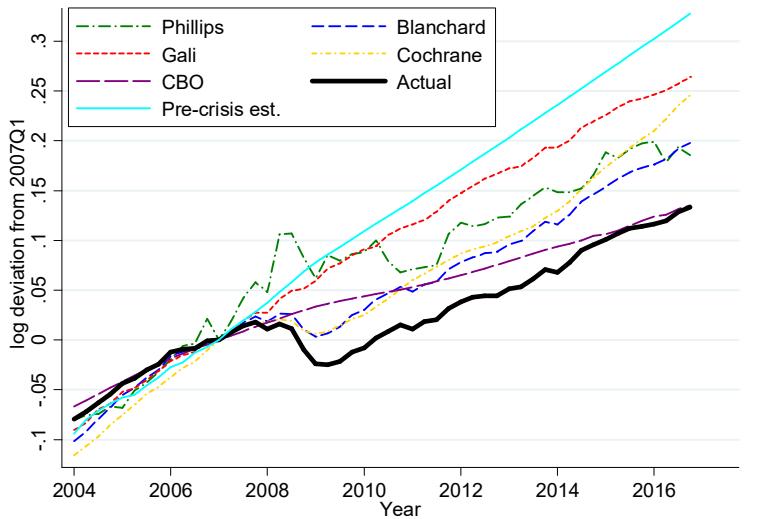
Panel B: Revisions using Gali (1999)



Panel C: Revisions using Cochrane (1994)



Panel D: Combined Estimates



Notes: Panels A, B and C plot the real-time estimates and forecasts of potential GDP following Blanchard and Quah (1989) in Panel A, Gali (1999) in Panel B and Cochrane (1994) in Panel C for different vintages of available data. Panel D 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 (“Phillips”), the CBO estimates of 2017 (“CBO”) and 2007 (“Pre-crisis est.”). In each panel, “Actual” denotes the path of Real GDP. See section 5.2 for details.

Table 1. Comparison of IMF, OECD and Consensus Economics.

	Institution and output measure		
	IMF, potential output growth rate (nowcast)	OECD, potential output growth rate (nowcast)	Consensus Economics, 6-10 year ahead forecast for actual output growth rates
Observations	607	1358	581
Mean	1.64	2.30	2.22
St. Deviation	1.10	1.25	0.54
Correlation			
IMF	1.00		
OECD	0.87	1.00	
Consensus Economics	0.72	0.78	1.00

Notes: The table reports moments of measures of potential output from the IMF and OECD across countries described in Appendix Table 1, as well as moments of forecasted growth rates of GDP 6-10 years ahead from Consensus Economics. See section 2.6 for details.

Table 2. Predictability of Revisions in Estimates of Potential GDP.

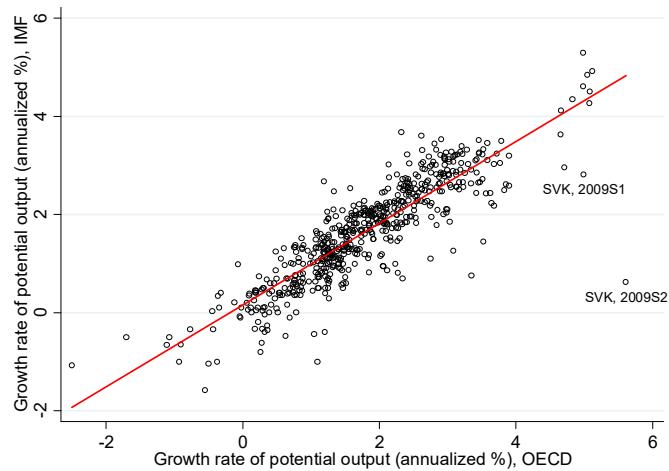
Dependent variable: $(\log Y_{t t}^* - \log Y_{t t-1}^*)$	Source				
	CBO	Greenbook	OECD	IMF	Consensus Economics
	(1)	(2)	(3)	(4)	(5)
$(\log Y_{t-1 t-1}^* - \log Y_{t-1 t-2}^*)$	0.204 (0.132)	0.294*** (0.086)	-0.066 (0.040)	-0.154*** (0.044)	-0.355*** (0.045)
Observations	42	96	1,282	548	566
R-squared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

Notes: The table presents regressions of the revision in estimates of potential GDP on the previous revision in estimate of potential GDP (equation 1). Newey-West standard errors are in parentheses. “Source” indicates where estimates of potential output come from: Congressional Budget Office (CBO), Greenbooks of the Federal Reserve Board (FED), the Organization for Economic Cooperation and Development (OECD), the International Monetary Fund (IMF) or Consensus Economics (CE). For the latter, revisions are for growth rate of GDP at horizons of 6-10 years. Columns (3)-(5) are across countries and include time and country fixed effects. Within R2 is reported for columns (3)-(5).

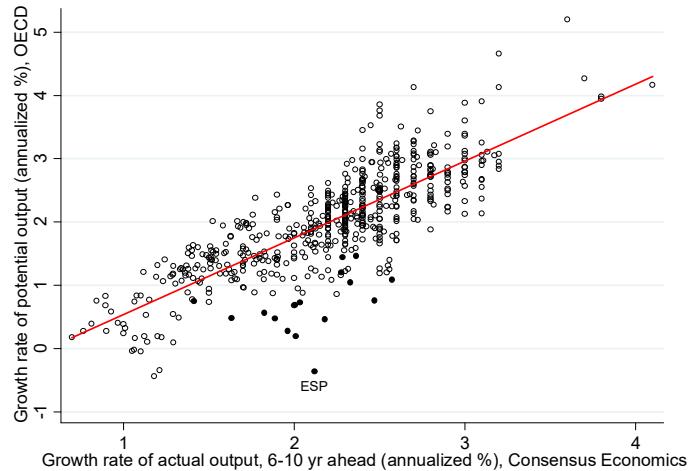
Appendix

Appendix Figure 1: Comparison of IMF and OECD estimates (nowcast) for potential output growth rate with forecasted long-term growth for actual output in Consensus Economics.

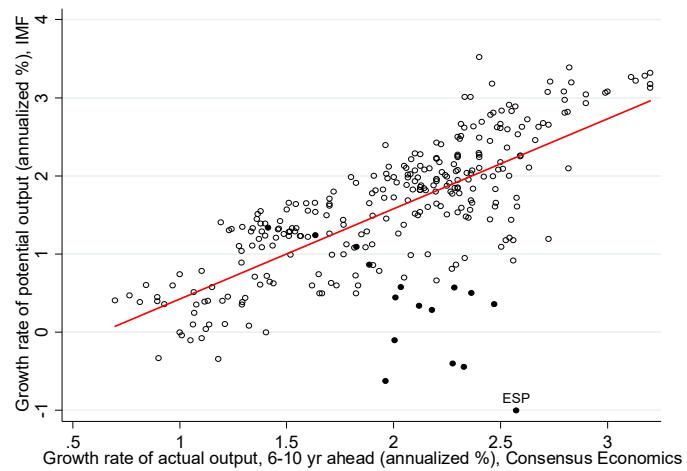
Panel A. IMF vs OECD



Panel B. OECD vs. Consensus Economics

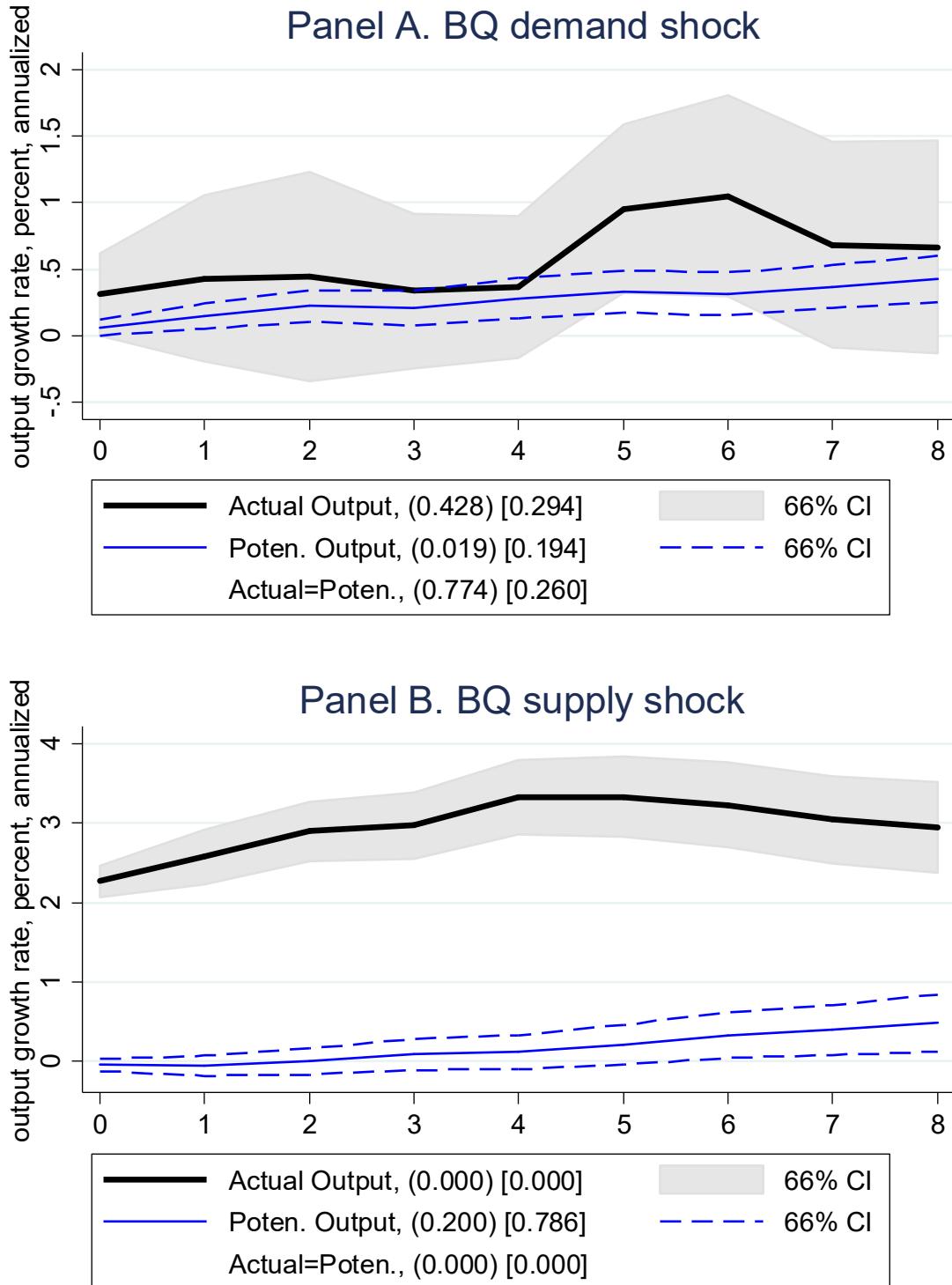


Panel C: IMF vs Consensus Economics



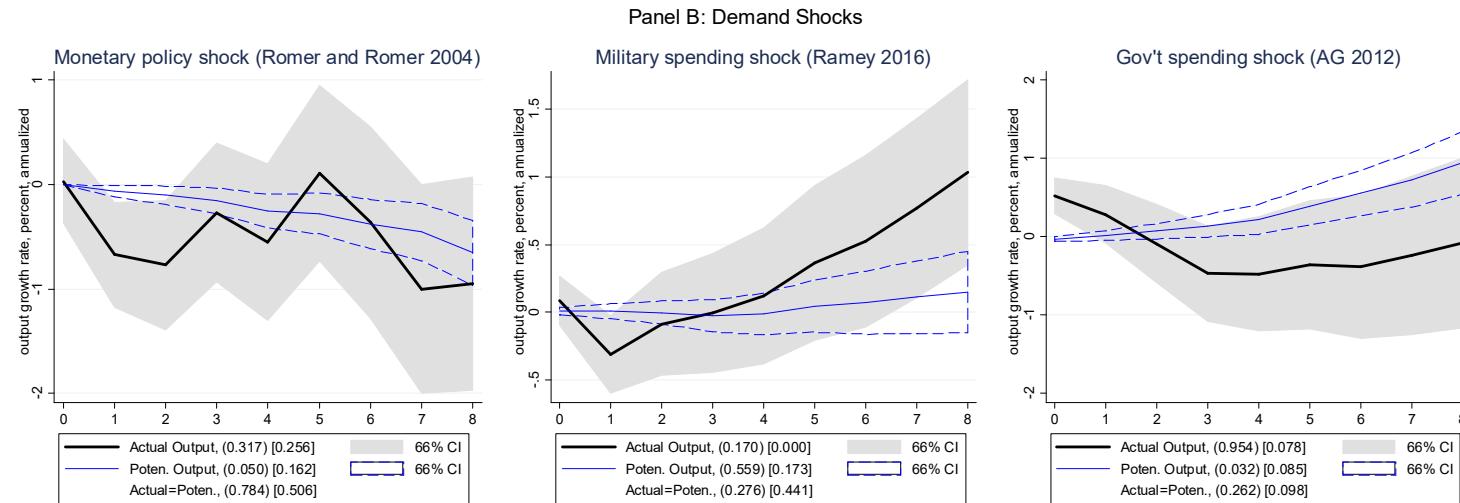
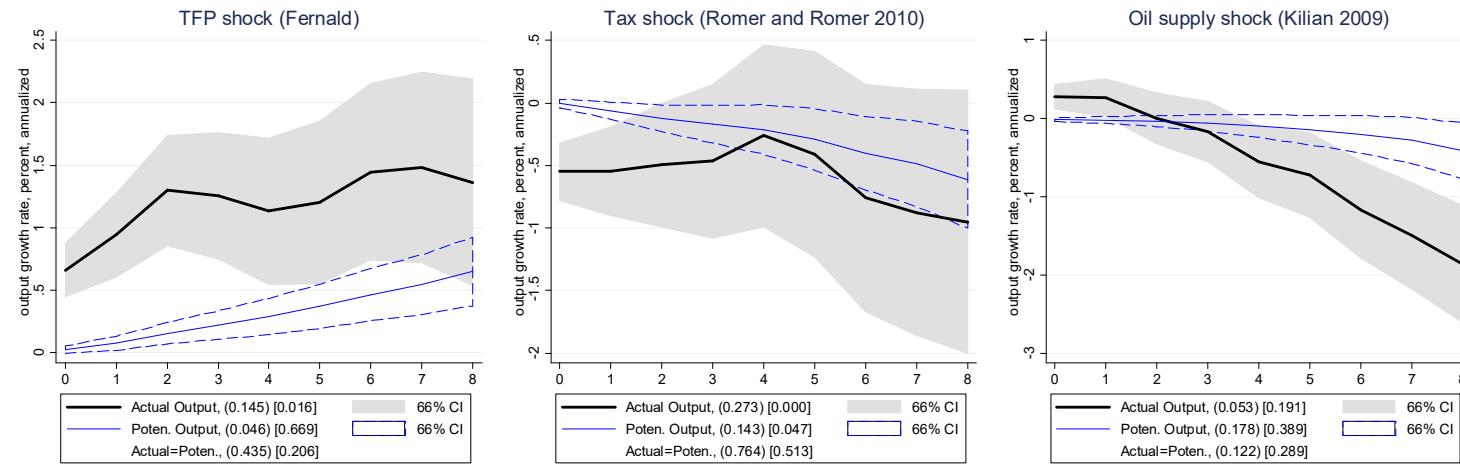
Notes: Filled markers in Panels B and C show observations for Spain in the 2009-2016 period.

Appendix Figure 2: Responses to BQ Identified Supply and Demand Shocks



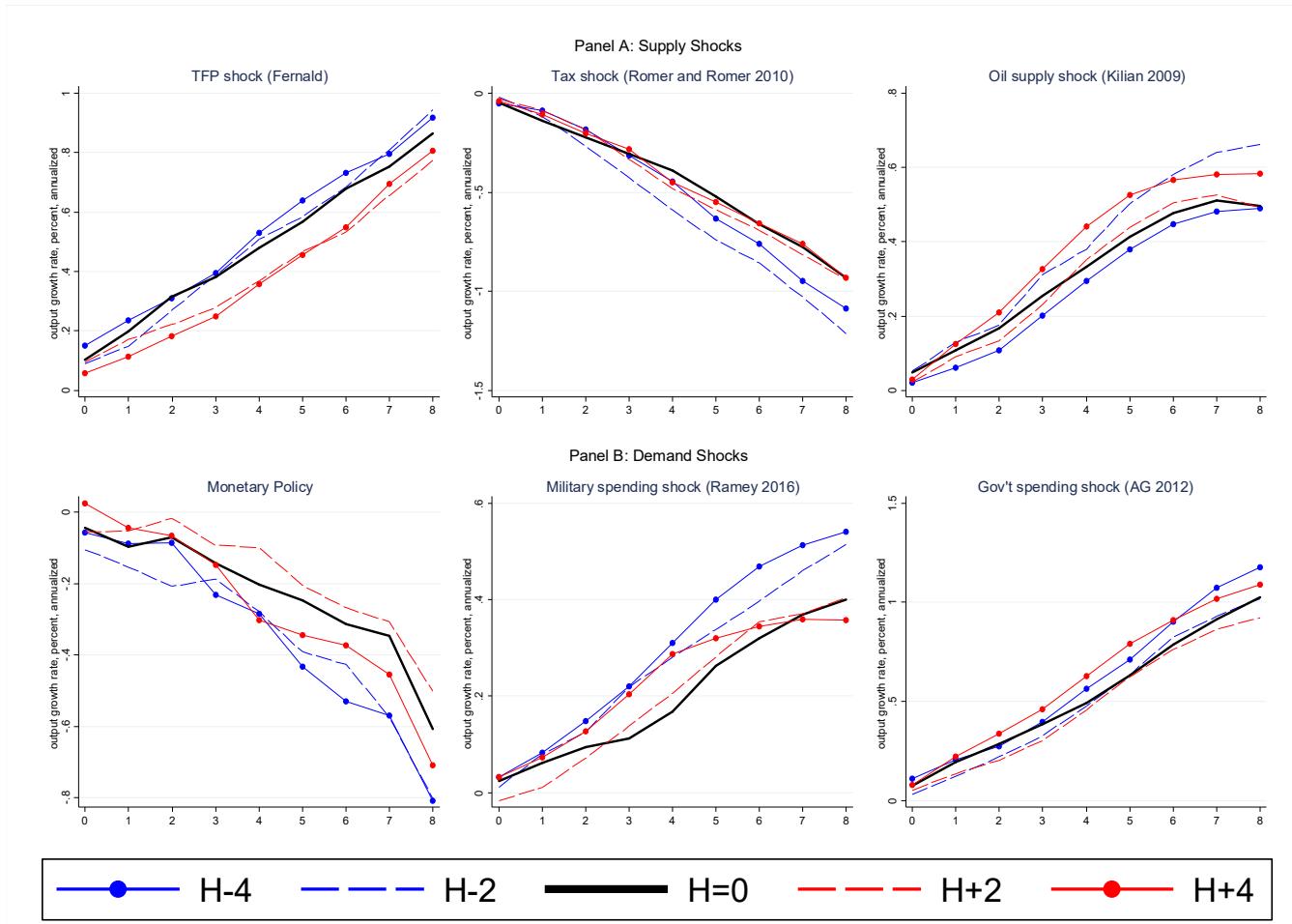
Notes: The figure reports impulse response functions (IRFs) estimated using equation (2) and (3). The estimation sample covers the benchmark time period for Greenbook forecasts. “Supply” and “Demand” shocks are identified as in Blanchard and Quah (1989). In parentheses we report the p-value for a test of whether the IRF of actual (potential) output is different from zero at the max horizon (8 quarters), while in square brackets we show the p-value for a test of whether the path of the IRF of actual (potential) output is different from zero over the entire duration of the IRF. The last row of the legend reports p-values for a test of equality of IRFs of actual and potential output at the max horizon (parentheses) and a test of equality of the paths of the responses for actual and potential output are equal across horizons.

Appendix Figure 3: Responses of Output and Greenbook Estimates of Potential Output in U.S. to Shocks: ADL specification.
 Panel A: Supply Shocks



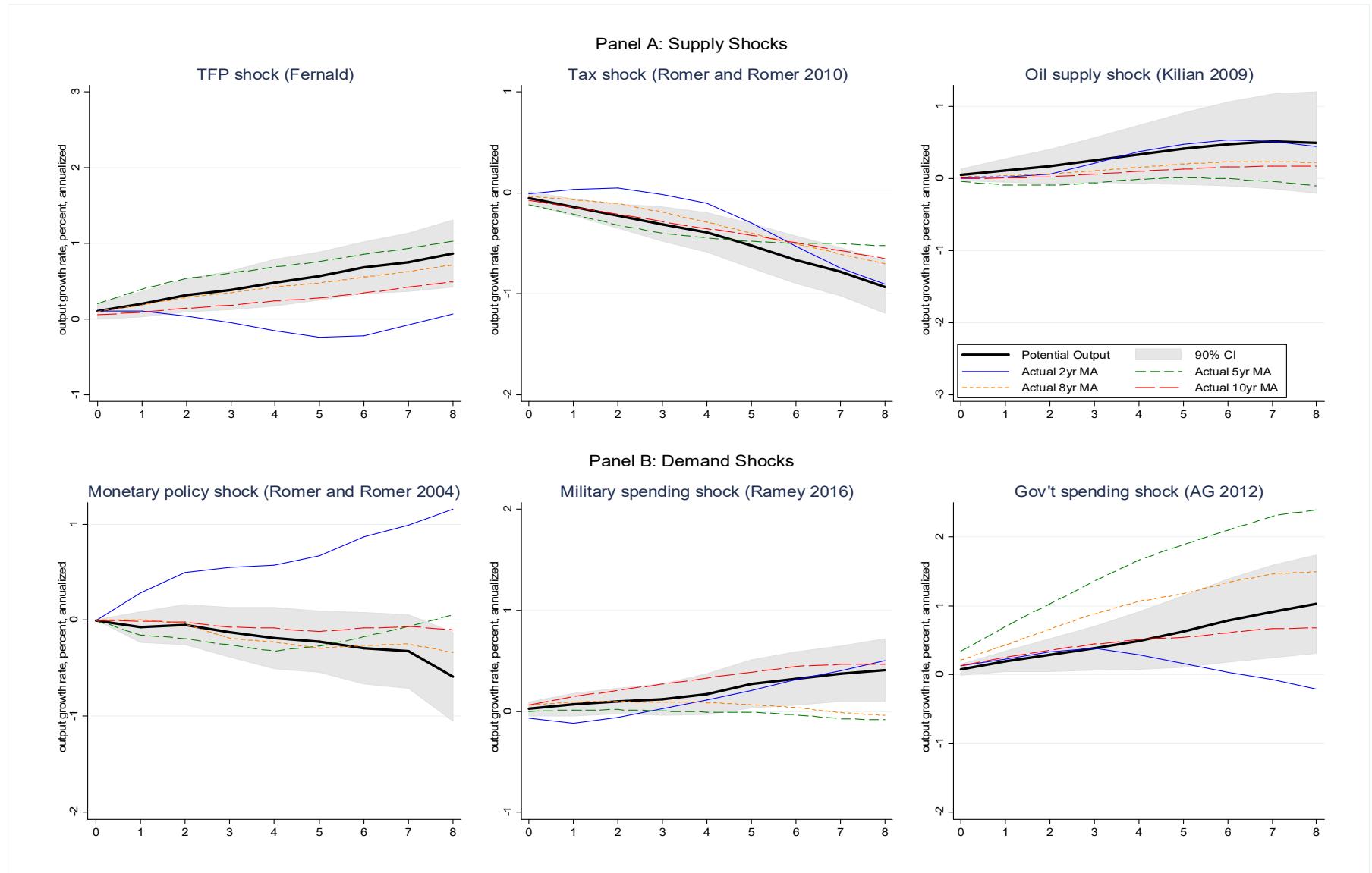
Notes: The figure reports impulse response functions (IRFs) estimated using equation (4), which is an auto-distributed lag specification. The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) available at the Federal Reserve Bank of Philadelphia. In parentheses we report the p-value for a test of whether the IRF of actual (potential) output is different from zero at the max horizon (8 quarters), while in square brackets we show the p-value for a test of whether the path of the IRF of actual (potential) output is different from zero over the entire duration of the IRF. The last row of the legend reports p-values for a test of equality of IRFs of actual and potential output at the max horizon (parentheses) and a test of equality of the paths of the responses for actual and potential output are equal across horizons.

Appendix Figure 4: Responses of Backcasts and Forecasts of Potential Output.



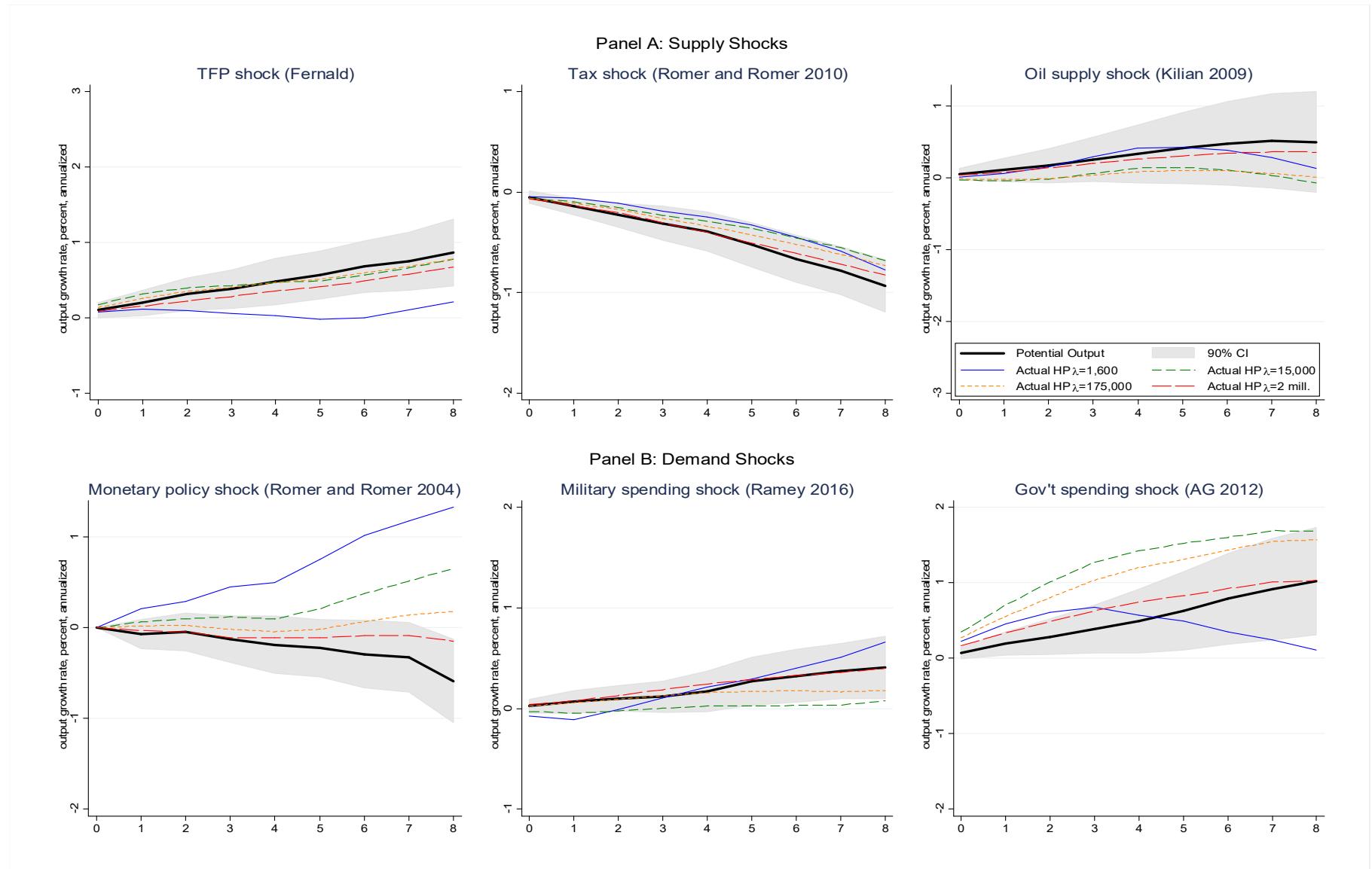
Notes: the figure shows impulse responses of horizon $H + k$ growth rate of potential output to structural shocks. $k > 0$ corresponds to forecasts, $k < 0$ correspond to backcasts, $k = 0$ is the nowcast (which corresponds to the results reported in Figure 6). All data are from Greenbooks.

Appendix Figure 5: Responses of Moving-Averages of Real-Time U.S. Output to Shocks.



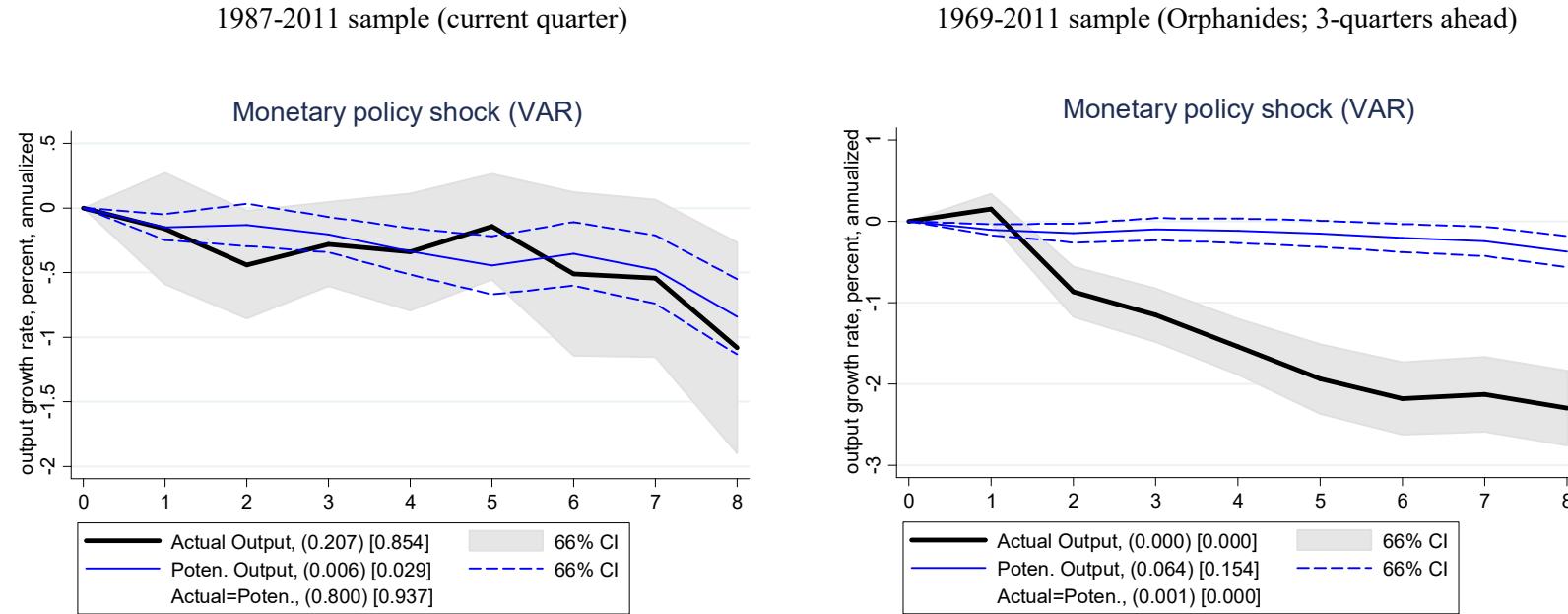
Notes: The figure reports impulse response functions (IRFs) estimated using equations (2) and (3). The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) available at the Federal Reserve Bank of Philadelphia.

Appendix Figure 6: Responses of HP-filters of Real-Time U.S. Output to Shocks.



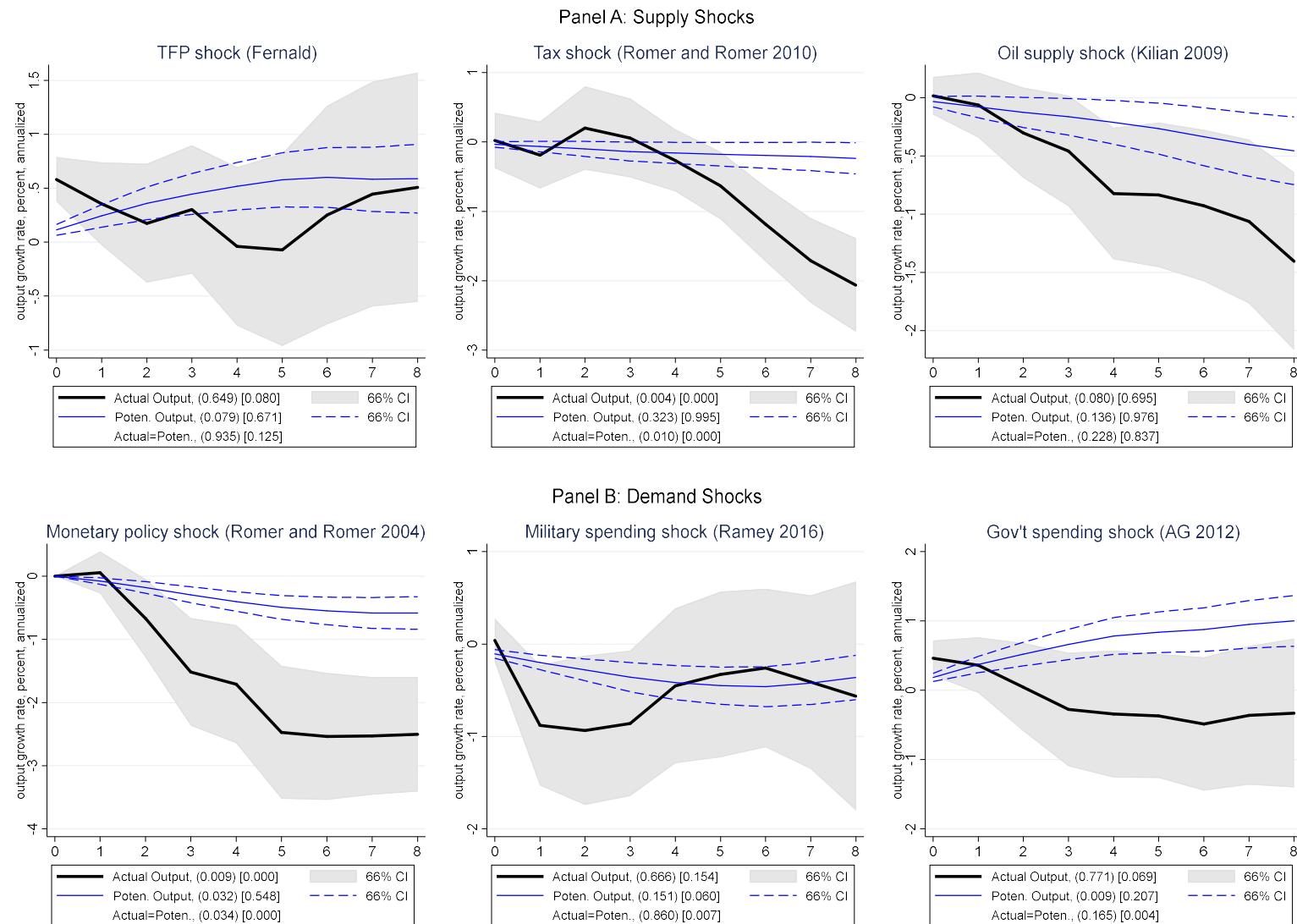
Notes: The figure reports impulse response functions (IRFs) estimated using equations (2) and (3). The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) available at the Federal Reserve Bank of Philadelphia.

Appendix Figure 7: Robustness of Responses to Identification of Monetary Shocks.



Notes: The figure reports impulse response functions (IRFs) estimated using equations (2) and (3). The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) available at the Federal Reserve Bank of Philadelphia (left panel) and the extended measure of potential GDP from Orphanides (2004) in right panel. Monetary shocks are identified from a trivariate VAR(4) using Cholesky restrictions.

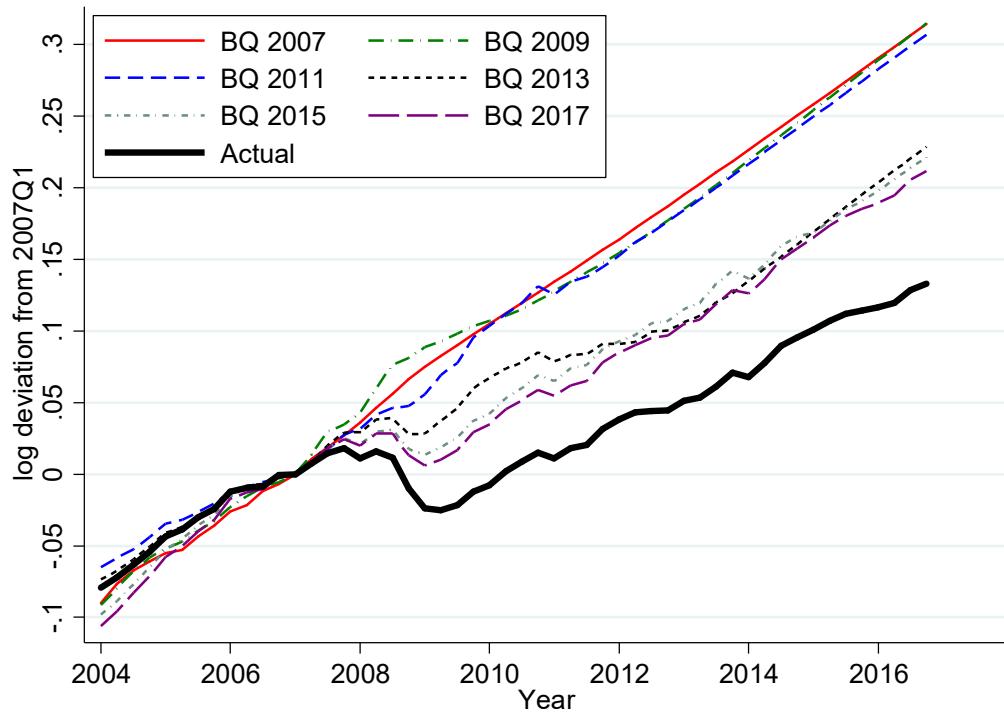
Appendix Figure 8: Responses of Final CBO Estimates of Potential to Economic Shocks



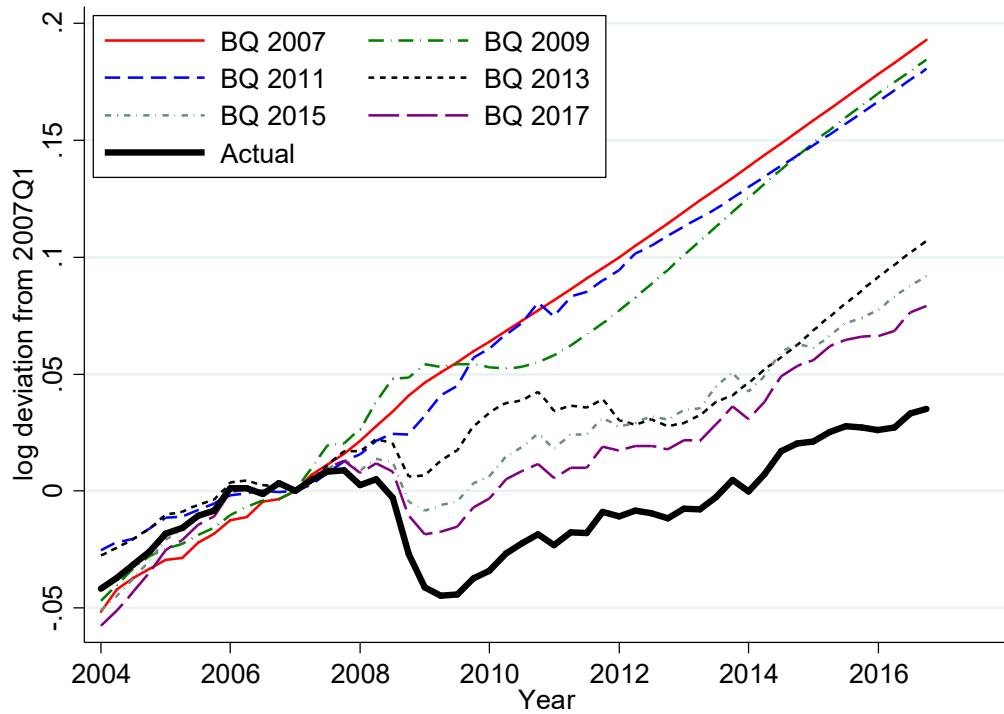
Notes: The figure reports impulse response functions (IRFs) estimated using equations (2) and (3). The estimation is identical to the baseline, except using final (2017) CBO estimates of potential GDP instead of real-time estimates.

Appendix Figure 9: Robustness of Responses to BQ Estimates of Monetary Shocks.

Panel A. Use Growth Rate of Output over the 1977-2007 Period to Measure Trend Growth



Panel B. Normalize Output by Non-Institutional Civilian Population



Notes: the figure shows the sensitivity of BQ analysis to alternative assumptions. Panel A uses the growth rate of output over the 1977-2007 period rather than the 1947-2007 period to measure trend growth of real GDP. Panel B normalized real GDP by non-institutional civilian population and use the 1947-2007 period to compute trend growth for the normalized variable.

Appendix Table 1. *Data coverage for cross-country analysis.*

Country	Prod. Shock	Oil Shock	Monetary Shock	Fiscal Shock	Actual IMF	Potential IMF	Actual OECD	Potential OECD	Actual C.E.
Australia	1981-2018	1980-2016	1983-2016	1998-2014	2003-2016	2003-2016	1986-2016	1989-2016	No data
Austria	No data	1980-2016	1989-2016	1998-2014	2003-2016	2003-2016	1986-2016	1989-2016	No data
Belgium	1981-2018	1980-2016	1984-2016	1998-2013	2003-2016	2003-2016	1986-2016	1989-2016	No data
Canada	1981-2018	1980-2016	1994-2016	1987-2014	2003-2016	2003-2016	1986-2016	1989-2016	1989-2016
Switzerland	No data	1980-2016	1994-2016	1998-2014	No data	No data	1986-2016	1989-2016	1998-2016
Cyprus	No data	1980-2015	2001-2016	No data	2003-2016	2009-2016	No data	No data	No data
Czech Republic	1994-2018	1990-2016	1996-2016	1998-2009	No data	No data	1996-2016	2005-2016	No data
Germany	1992-2018	1980-2016	1994-2016	1987-2014	2003-2016	2003-2016	1986-2016	1989-2016	1989-2016
Denmark	No data	1980-2016	1984-2016	1998-2010	2003-2016	2009-2016	1986-2016	1989-2016	No data
Spain	No data	1980-2016	1987-2016	1998-2012	2003-2016	2003-2016	1986-2016	1989-2016	1995-2016
Estonia	1996-2018	1990-2016	1995-2016	2010-2014	2003-2016	2012-2016	2008-2016	2011-2016	No data
Finland	1981-2018	1980-2016	1989-2016	1998-2014	2003-2016	2003-2016	1986-2016	1989-2016	No data
France	1981-2018	1980-2016	1983-2016	1987-2014	2003-2016	2003-2016	1986-2016	1989-2016	1989-2016
United Kingdom	1981-2018	1980-2016	1990-2016	1987-2014	2003-2016	2003-2016	1986-2016	1989-2016	1989-2016
Greece	No data	1980-2016	No data	1998-2001	2003-2016	2009-2016	1986-2016	1989-2016	No data
Hungary	No data	1980-2016	2002-2016	1998-2003	No data	No data	1996-2016	2005-2016	No data
Ireland	1991-2018	1980-2016	2000-2016	1998-2014	2003-2016	2003-2016	1996-2016	1996-2016	No data
Iceland	1981-2018	1980-2016	1999-2016	1998-2014	No data	No data	1986-2016	2000-2016	No data
Italy	1981-2018	1980-2016	1984-2016	1987-2014	2003-2016	2003-2016	1986-2016	1989-2016	1989-2016
Japan	1981-2018	1980-2016	1994-2016	1987-2014	2003-2016	2003-2016	1986-2016	1989-2016	1989-2016
Korea	1981-2018	1980-2016	1994-2016	1999-2014	2003-2016	2012-2016	1997-2016	2005-2016	No data
Luxembourg	1986-2018	1980-2016	1997-2016	1998-2014	2003-2016	2012-2016	1986-2016	2005-2016	No data
Malta	No data	1980-2015	No data	No data	2003-2016	2009-2016	No data	No data	No data
Netherlands	1981-2018	1980-2016	1984-2016	1998-2014	2003-2016	2003-2016	1986-2016	1989-2016	1995-2016
Norway	1981-2018	1980-2016	1981-2016	1998-2014	2003-2016	2003-2016	1986-2016	1989-2016	1998-2016
New Zealand	1990-2018	1980-2016	1987-2016	1998-2014	2003-2016	2003-2016	1986-2016	1989-2016	No data
Poland	No data	1980-2016	1997-2015	1998-2011	No data	No data	1996-2016	2005-2016	No data
Portugal	1981-2018	1980-2016	1993-2016	1998-2014	2003-2016	2003-2016	1986-2016	1994-2016	No data
Slovak Republic	No data	1980-2016	2001-2016	2008-2009	2003-2016	2009-2016	2000-2016	2005-2016	No data
Slovenia	No data	1992-2016	1997-2016	2014-2014	2003-2016	2009-2016	2008-2016	2010-2016	No data
Sweden	1981-2018	1980-2016	1984-2016	1998-2014	2003-2016	2003-2016	1986-2016	1989-2016	1995-2016
Turkey	No data	1980-2016	2001-2016	1998-2002	No data	No data	1986-2016	2005-2016	No data
United States	1981-2018	1980-2016	1981-2016	1987-2014	2003-2016	2003-2016	1986-2016	1989-2016	1989-2016

Notes: The table describes time periods for which shocks and measures of potential output are available for each country and source of data. “C.E.” are forecasts of 6-10 year ahead GDP growth. See section 2 for descriptions of measures of potential GDP, and sections 3 and 4 for details on construction of shocks.

Appendix Table 2. P-values for tests for U.S. data

Shocks	Measure of actual output		Potential output		Equality of IRFs for measure of actual and potential output	
	IRF is equal to zero pointwise	IRF is zero at the max horizon	IRF is equal to zero pointwise	IRF is zero at the max horizon	pointwise	at the max horizon
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Greenbook, 1987-2011, Measure of actual = actual						
TFP shock	0.020	0.296	0.126	0.001	0.174	0.962
Government spending shock, (AG 2012)	0.065	0.922	0.093	0.017	0.030	0.336
Tax shock (RR 2010)	0.002	0.106	0.000	0.000	0.001	0.983
Military spending shock (Ramey 2016)	0.000	0.204	0.183	0.029	0.000	0.506
Oil price shock (Kilian 2009)	0.018	0.012	0.894	0.242	0.038	0.002
Monetary policy shock (RR 2004)	0.409	0.533	0.455	0.035	0.446	0.788
Panel B. Greenbook, 1969-2011, Measure of actual = actual						
TFP shock	0.030	0.687	0.452	0.067	0.048	0.930
Government spending shock, (AG 2012)	0.012	0.919	0.901	0.163	0.019	0.479
Tax shock (RR 2010)	0.000	0.004	0.548	0.027	0.001	0.070
Military spending shock (Ramey 2016)	0.107	0.728	0.002	0.000	0.264	0.450
Oil price shock (Kilian 2009)	0.409	0.067	0.007	0.000	0.136	0.002
Monetary policy shock (RR 2004)	0.000	0.018	0.001	0.005	0.000	0.100
Panel C1. Greenbook, 1987-2011, Measure of actual = 5yr MA of last vintage of actual						
TFP shock	0.441	0.016	0.126	0.001	0.991	0.935
Government spending shock, (AG 2012)	0.041	0.001	0.093	0.017	0.408	0.069
Tax shock (RR 2010)	0.977	0.868	0.000	0.000	0.096	0.077
Military spending shock (Ramey 2016)	0.955	0.218	0.183	0.029	0.539	0.020
Oil price shock (Kilian 2009)	0.967	0.296	0.894	0.242	0.236	0.002
Monetary policy shock (RR 2004)	0.313	0.461	0.455	0.035	0.000	0.012
Panel C2. Greenbook, 1987-2011, Measure of actual = 5yr MA of real time actual						
TFP shock	0.488	0.008	0.126	0.001	0.980	0.567
Government spending shock, (AG 2012)	0.004	0.000	0.093	0.017	0.079	0.011
Tax shock (RR 2010)	0.973	0.285	0.000	0.000	0.334	0.363
Military spending shock (Ramey 2016)	0.999	0.794	0.183	0.029	0.776	0.116
Oil price shock (Kilian 2009)	0.993	0.893	0.894	0.242	0.953	0.140
Monetary policy shock (RR 2004)	0.806	0.884	0.455	0.035	0.000	0.008
Panel C3. Greenbook, 1987-2011, Measure of actual = HP of real time actual						
TFP shock	0.514	0.010	0.126	0.001	0.951	0.266
Government spending shock, (AG 2012)	0.205	0.010	0.093	0.017	0.198	0.986
Tax shock (RR 2010)	0.089	0.001	0.000	0.000	0.344	0.567
Military spending shock (Ramey 2016)	0.779	0.078	0.183	0.029	0.063	0.963
Oil price shock (Kilian 2009)	0.998	0.419	0.894	0.242	0.910	0.470
Monetary policy shock (RR 2004)	0.998	0.640	0.455	0.035	0.000	0.001
Panel D. CBO, 1991-2011, Measure of actual = actual						
TFP shock	0.250	0.041	0.000	0.001	0.916	0.843
Government spending shock, (AG 2012)	0.290	0.141	0.017	0.001	0.360	0.922
Tax shock (RR 2010)	0.000	0.024	0.000	0.001	0.000	0.984
Military spending shock (Ramey 2016)	0.000	0.006	0.000	0.000	0.382	0.636
Oil price shock (Kilian 2009)	0.017	0.227	0.959	0.503	0.036	0.031
Monetary policy shock (RR 2004)	0.994	0.922	0.720	0.844	0.900	0.959

Notes: The table reports p-values for responses of actual GDP (columns 1-2) or estimates of potential GDP (columns 3-4) in response to shocks. Column 1 tests null that actual GDP is always zero in IRFs, column 2 tests null that its response is zero at the max horizon of IRFs. Columns 3 and 4 are equivalent but for responses of the estimates of potential GDP. Column 5 tests the null that the IRFs of actual GDP and estimated potential are the same at all horizons while column 6 tests the null they are the same at the final horizon. Panels A and C (1, 2 and 3) use the same measure of potential GDP (Greenbook 1987-2001); what changes between these panels is the measure of actual GDP (panel A uses the last vintage of actual output, panel C1 uses a 5 year moving average of the last vintage of actual output, panel C2 uses a 5 year moving average of actual output in real time and panel C3 uses an actual output in real time filtered with the Hodrick and Prescott method).

Appendix Table 3. P-values for tests for international data

Shocks	Measure of actual output		Potential output		Equality of IRFs for measure of actual and potential output	
	IRF is equal to zero pointwise	IRF is zero at the max horizon	IRF is equal to zero pointwise	IRF is zero at the max horizon	pointwise	at the max horizon
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. IMF, Measure of actual = actual						
TFP shock	0.000	0.000	0.046	0.011	0.000	0.000
Oil price shock (Kilian 2009)	0.001	0.098	0.007	0.101	0.008	0.171
Monetary policy shock (VAR)	0.000	0.123	0.002	0.036	0.000	0.128
Government spending shock, (AG 2012)	0.033	0.364	0.001	0.002	0.086	0.825
Panel B. IMF, Measure of actual = HP of real time actual						
TFP shock	0.289	0.128	0.046	0.011	0.301	0.272
Oil price shock (Kilian 2009)	0.004	0.002	0.007	0.101	0.000	0.000
Monetary policy shock (VAR)	0.003	0.000	0.002	0.036	0.163	0.433
Government spending shock, (AG 2012)	0.023	0.000	0.001	0.002	0.139	0.430
Panel C. OECD, Measure of actual = actual						
TFP shock	0.000	0.000	0.003	0.000	0.000	0.000
Oil price shock (Kilian 2009)	0.061	0.445	0.338	0.081	0.117	0.955
Monetary policy shock (VAR)	0.080	0.023	0.470	0.070	0.289	0.118
Government spending shock, (AG 2012)	0.002	0.081	0.000	0.000	0.002	0.583
Panel D. OECD, Measure of actual = HP of real time actual						
TFP shock	0.001	0.000	0.003	0.000	0.000	0.004
Oil price shock (Kilian 2009)	0.578	0.313	0.338	0.081	0.173	0.361
Monetary policy shock (VAR)	0.044	0.001	0.470	0.070	0.052	0.001
Government spending shock, (AG 2012)	0.000	0.000	0.000	0.000	0.577	0.486
Panel E. Consensus Economics, Measure of actual = actual						
TFP shock	0.000	0.000	0.226	0.019	0.000	0.000
Oil price shock (Kilian 2009)	0.116	0.707	0.020	0.003	0.065	0.370
Monetary policy shock (VAR)	0.001	0.000	0.938	0.418	0.027	0.001
Government spending shock, (AG 2012)	0.025	0.986	0.074	0.002	0.018	0.583
Panel F. Consensus Economics, Measure of actual = HP of real time actual						
TFP shock	0.043	0.001	0.226	0.019	0.051	0.001
Oil price shock (Kilian 2009)	0.511	0.075	0.020	0.003	0.794	0.842
Monetary policy shock (VAR)	0.073	0.070	0.938	0.418	0.074	0.057
Government spending shock, (AG 2012)	0.006	0.000	0.074	0.002	0.018	0.000

Notes: The table reports p-values for different statistics of responses of actual GDP (columns 1-2) or estimates of potential GDP (columns 3-4) in response to shocks listed in the table using different measures of potential GDP. Column 1 tests null that actual GDP is always zero in IRFs while column 2 tests null that its response is zero at the maximum horizon of IRFs. Columns 3 and 4 are equivalent but for responses of the estimates of potential GDP. Column 5 tests the null that the IRFs of actual GDP and estimated potential are the same at all horizons while column 6 tests the null they are the same at the final horizon. See section 4 for details. Notice also that the measure of potential output is the same in panels A and B, in panels C and D and in panels E and F, what differs between these pairs is that the first uses the last vintage of actual output as a measure of actual output while the second uses real time actual output filtered with an HP filter with $\lambda = 800$.