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: June 21st, 2017

Abstract: The fact that most of the persistent 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. Using a variety of estimates of potential output for the U.S. and other countries, we show that these estimates respond gradually not only to supply-side shocks but also respond to demand shocks that have only transitory effects on output. Observing a revision in measures of potential output therefore says little about whether concurrent changes in actual output are likely to be permanent or not. In contrast, some structural VAR methodologies can avoid these shortcomings, even in real-time. This approach points toward a more limited decline in potential output following the Great Recession.

JEL: E2, E3
Keywords: Potential output, output persistence, secular stagnation.

This 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 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 now nearly 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. In light of these revisions, a number of commentators (e.g. Summers 2014, 2016, Krugman 2014) have suggested that the U.S. economy has entered a period of secular stagnation. If correct, this interpretation poses many, seemingly insurmountable challenges for policymakers in the current environment of ultra-low interest rates and limited fiscal capacity to stimulate economic activity.

However, before we take these dynamics in the estimates of potential output and ring the alarm bell, we should understand the properties of estimates of potential output and what determines revisions of the estimates. In this paper, we focus on how real-time estimates of actual and potential output respond to different economic shocks in the U.S. as well as across a wide range of countries. Using a variety of sources that estimate 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 the gradual pace of adjustment to shocks and the inability of the estimates 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). We show that estimates of potential output for a country are highly correlated across organizations. 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.

As a result, 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, we find that 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 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 of these estimates correctly capturing 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 cyclical 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, whether it be for the U.S. or for a broader cross-section of countries. 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 real 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 sources underlying changes in GDP, be they monetary, technology, etc. Thus, we can rationalize why one can observe a gradual response to any economic shock, even those that have only transitory effects on GDP and that should in principle be stripped out of estimates of potential GDP.

We argue that some of the puzzling reactions of potential output to identified shocks may be addressed using tools developed in previous studies. Specifically, we show that the Blanchard and Quah (1989) approach to identify supply and demand shocks can be a useful element in a statistical framework that generates potential output estimates 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 different interpretation for changes in U.S. potential output following the Great Recession. Our estimates imply that U.S. output remains almost 10 percentage points below potential output, leaving ample room for policymakers to close the gap through demand-side policies if they so chose to.

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 Benigno and Fornaro (2017)), empirical evidence on it remains scant. 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 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 cyclical 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 different real-time estimates of potential GDP can in short samples depend on 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.

Finally, by comparing the 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), for example, 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 an example of how estimates of potential output can be improved. Section 6 concludes.

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

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 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).

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). This approach is 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.

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.

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.

Forecasts 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 of the course of the Great Recession, the CBO had similarly made 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 time period.

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 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 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 and inflation (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 output gap reported in Greenbooks since 1987.[1] 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.

[1] 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.

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 also 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 during this period were less concerned about inflation than they normally would have been given the low unemployment rates of this period (see Gorodnichenko and Shapiro 2007).

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 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 measure potential output is methodologically similar to measures reported in Greenbooks, in the sense that they use a combination of the 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 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 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, doesn't 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 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 2 documents some basic moments for estimates of potential output growth rate (nowcasts) produced by the IMF and OECD as well as forecasted long-term actual output growth rate from Consensus Economics. These series are highly correlated and generally have similar moments. 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.

Figure 2 shows that these strong correlations are not driven by outliers and that large differences across sources tend to be concentrated in a handful of countries and periods. For example, the largest difference between the IMF and OECD estimates of potential output growth rate happens for Slovakia in 2009, during which GDP fell sharply and the IMF reduced its estimates of potential GDP growth while the OECD did not. In a similar spirit, the IMF and OECD estimated a low growth rate of potential output for Spain in the post-Great Recession period while forecasters in Consensus Economics maintained their predictions of a relatively fast long-term growth rate of actual output for Spain during the same period.

Figure 3 illustrates that this strong correlation across series is not restricted to differences in growth rates across countries. Time series for the growth rate for 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 4 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 5-year moving average of real-time GDP and an HP-filter (λ=2,000,000) 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 3 and 4 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:

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

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 3) 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. In both cases, we find that including dummies for the large revisions associated with the September 11th attacks and the Great Recession eliminates any clear predictability in the revisions of estimates of potential GDP. Similar results obtain across countries using the OECD and IMF estimates of potential GDP, with coefficients on past OECD revisions being not different from zero while that on past IMF revisions being positive but small. In contrast, we find a negative predictability in Consensus Economics revisions of long-run growth rates.

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.[2]

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 Sims 2016), 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). 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 they appear to have permanent effects on output, and therefore should be captured by estimates of potential GDP.

We consider three identified demand shocks, all related to policy. The first are monetary policy shocks. For the 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.[3]

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

[2] 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, and we define demand shocks as those whose real effects appear to be transitory and therefore should not affect long-run forecasts of potential output. Because the units of these shocks vary, we normalize all shocks to be mean zero and have unit variance.

[3] 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 3.

these measures from one quarter to the next therefore reflect changes either 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 much more than just military spending.

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

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

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

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. This moving-average specification is similar to the direct projection method in Jorda (2005). A key advantage of this approach 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. 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.[4]

We plot the responses of actual output to each type of shock in Figure 5. 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.

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

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 are also generally delayed).[5] 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.

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

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

where $\Delta \log Y^*_{t|t}$ is the (nowcast) estimated growth in potential in quarter *t* given information in quarter *t* at an annualized rate. 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.

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, 2015). 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^{pot}) in the long run, as estimates of potential are progressively increased while output itself is falling. Policymakers facing a tradeoff between

[5] 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.

stabilizing inflation (which rises after a 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.[6] 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 this 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.

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:

[6] 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.

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

Using $J = 4$ and $K = 8$. Results are presented in Figure 6. 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 7, 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.[7] Hence, the baseline results are not specific to the period since 1987.

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 8, we reproduce our results using estimates of potential GDP from the Congressional Budget Office. One advantage of CBO estimates is they are available at different 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 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.

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

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

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 4, 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 moving-average of recent GDP outcomes.[8] 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 alternative statistical approaches to estimating potential GDP. In the first case, we apply a one-sided HP-filter 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 9. When using the HP-filtered series, we can very closely replicate the response of estimated potential GDP after every shock.[9] With the moving average, the fit is not as close, and the response goes in the wrong direction after news about military spending and oil supply shocks. 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 4 with an HP-filtered series, suggests that

[8] This one-sided filter is implemented as follows. For a given quarter, "potential" output is calculated as the value of the HP-filter trend for the quarter given the first vintage of GDP data that cover the given quarter.

[9] 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.

Greenbook estimates of potential GDP incorporate little additional information relative to this purely statistical approach to estimating potential GDP.[10] 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 Greenbook estimates of potential GDP is *not* informative about whether the associated declines in actual GDP are likely to be sustained or not.

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). 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.

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

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:

$$\Delta \log Y_{j,t|t} = \alpha_j + \gamma_t + \sum_{k=0}^{K} \phi_k \epsilon_{j,t-k} + error_{t,j} \tag{5}$$

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

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 frequency of real-time estimates of potential GDP by both the IMF and OECD.

[10] The best fit of HP-filtered series comes with very high values of λ (we use λ=2,000,000). This high value is consistent with a low pass filter that allows only low frequencies with periods of about 15 years and higher. Lower values do not replicate Greenbook measures of potential GDP as closely, as can be seen in Appendix Figure 1. Similarly with moving average measures, we can better replicate the dynamic response of Greenbook estimates of potential when averaging over periods (10-20 years) than over shorter horizons (3-5 years) as illustrated in Appendix Figure 2.

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 past changes in labor productivity as well as country and time fixed effects.[11] 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 time fixed effects.[12] 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.[13] 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 10 presents responses of both GDP and potential to each of the four shocks. 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 have mostly dissipated after two to three years.

The effects of these shocks on potential GDP are very consistent with those obtained for the U.S. with Greenbook and CBO forecasts. 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 eventually rise even though actual output falls in the long run. 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 highly transitory. Thus, we observe both the under-cyclicality after productivity shocks and over-cyclicality after demand shocks documented in the U.S., as well as the counter-cyclicality after oil supply shocks.

---

[11] 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.

[12] 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 has a lot more oil that it demands and can export a lot, for example Norway).

[13] 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.

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 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 11, 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. The one difference is in their response to oil supply shocks: IMF estimates of potential GDP decline (as does actual output) whereas those of the OECD (like the Greenbooks in the US) rise. 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 organization's estimates of potential GDP in a country possibly 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 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 12), 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, therefore displaying the same pattern as those of the IMF but not the OECD or the Greenbooks.

For comparison, we also plot the implied response of HP-filtered levels of output to the same shocks and countries. For all but oil supply 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). As was found with CBO forecasts at different horizons, long-run forecasts may be changed less rapidly than estimates of contemporaneous output gaps.

5 A Method to Estimate 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 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 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.

In this simple, proof-of-concept exercise, we follow Blanchard and Quah (1989) 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 a zero effect on the level of output in the long run. Note that these restrictions are broadly consistent with the responses of demand observed in Figure 5 and other results in the literature, namely that monetary and government spending shocks do not have permanent effects on output (e.g. Romer and Romer 2004, Ramey 2016). We then interpret predicted movements in output driven by supply-side shocks as capturing potential output.

Because Blanchard and Quah (1989) and others emphasize the importance of structural breaks, we use a rolling window of 120 quarters (results are similar when we use alternative rolling windows such as 80, 100, 140 or 160 quarters).

When applying the Blanchard and Quah (1989) approach, we use real-time data to ensure that our results are not driven by information not available to the econometrician. That is, in a particular quarter (say 1985Q1) 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), we estimate the SVAR with long run restriction using these series, we then perform the historical decomposition on this data to recover the component of the growth rate of actual output due to supply-side shocks, but from all the data available (a time series like 1960Q1-1985Q1), we keep only the data point that corresponds to the last quarter (1985Q1) and discard all the previous quarters. The next quarter's (1985Q2) historical decomposition data point is going to use vintages that were not available yet in 1985Q1, and the previous quarter's (1984Q4) historical decomposition data point used vintages that contained less information and stopped in 1984Q4. This approach therefore uses no more information than was available to agents in real-time, making our estimates comparable to real-time estimates of US potential GDP like those from the Greenbooks or the CBO.

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 13 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 Blanchard and Quah (1989) approach can therefore make progress toward resolving puzzles in the reaction of conventional estimates of potential output to identified shocks.

The different estimates of potential output coming from the Blanchard and Quah (1989) methodology are not innocuous in their policy implications relative to the estimates of organizations like the CBO. In Figure 14, we plot the real-time revisions in potential output from the Blanchard and Quah (1989) methodology. Like CBO 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 CBO estimates, the Blanchard and Quah (1989) approach points to a large difference and continuing gap between actual output and potential. In 2016, we estimate U.S. potential output to be nearly 10 log points higher than actual output, a difference which could potentially be closed through the use of demand side policies. Naturally, this is just a first step—in what will likely be many—to build a potential output framework consistent with theoretical insights. Further refinements such as external instruments, additional variables in the SVAR, and judgmental adjustments may be necessary. But the magnitude of the difference between our estimate of potential GDP and those of organizations like the CBO, combined with the potential policy implications of these differences, suggests that this is an area that is well worth pursuing further.

6 Conclusion

In the U.S. as well as across a wide range of countries, we find that private and public estimates of potential GDP respond gradually but systematically to all of the economic shocks that we consider and deviate little from what one would expect from simple univariate time series estimates of potential GDP. These results have several potential policy implications.

The first is that revisions in estimates of potential GDP tell us little about the underlying source of changes in GDP. While revisions in potential GDP are often interpreted as indicating permanent changes in the level of GDP, our results call for caution in adopting this interpretation. Even shocks that induce only transitory changes in income are associated with subsequent revisions in estimates of potential GDP. The fact that forecasters now attribute much of the decline in output across countries since the Great Recession to changes in potential GDP therefore tells us little about whether these changes in output are in fact likely to persist or whether they can be reversed through monetary or fiscal policies.

A second implication is that there is much work to be done to create better measures of potential GDP in real-time before policymakers rely on these too much. There are several methods that seem potentially underused. One is using additional macroeconomic variables and restrictions to better identify supply and demand shocks rather than relying on univariate processes. For example, we show that the Blanchard and Quah (1989) approach may provide a good starting point. A second possibility is to combine information from public estimates of potential GDP with private sector forecasts, as the latter appear somewhat more successful at isolating supply shocks from demand shocks. A third possibility is to avoid excessive use of model-averaging, or at least to avoid including simple approaches like HP-filters among the class of models used, since these mechanically induce movements in estimates of potential after cyclical demand-driven fluctuations. More generally, the absence of clear ways to precisely estimate potential output in real-time suggests that the practice of relying on “judgement” by professional economists should not be discontinued anytime soon.

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.

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.

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," Working Paper Series 2038, European Central Bank.

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.

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

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

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

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

Coibion, Olivier and Gorodnichenko, Yuriy, 2015. "Information Rigidity 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.

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

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

Resende, Carlos de, 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.

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.

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

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.

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).

Jordà, Òscar, 2005. "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review* 95(1): 161-182.

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.

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

Krugman, Paul (2014). “Four observations on secular stagnation.” in *Secular Stagnation: Facts, Causes and Cures,* Teulings, C. and Baldwin, R., eds., CEPR, London.

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.

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. 2001. “Monetary Policy Rules Based on Real-Time Data.” *American Economic Review* 91(4): 964-985.

Orphanides, Athanasios and van Norden, Simon, 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.

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

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

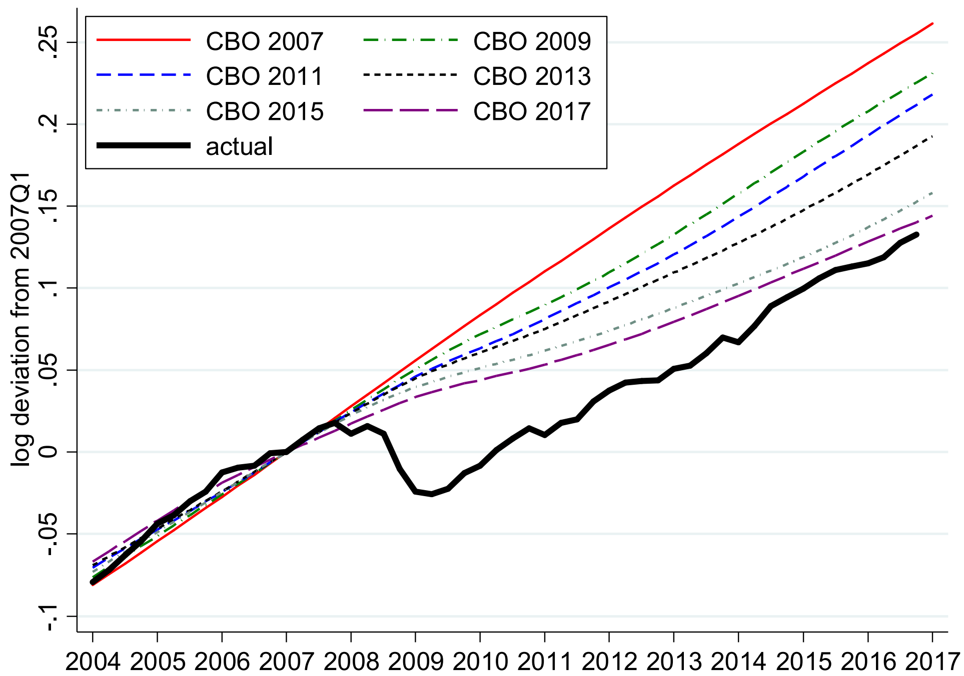
Sims, Eric, 2016. "Differences in Quarterly Utilization-Adjusted TFP by Vintage, with an Application to News Shocks." *NBER Working Paper*, 22154.

Summers, Lawrence, 2014. "U.S. Economic Prospects: Secular Stagnation, Hysteresis, and the Zero Lower Bound." *Business Economics* 49(2):65-73.

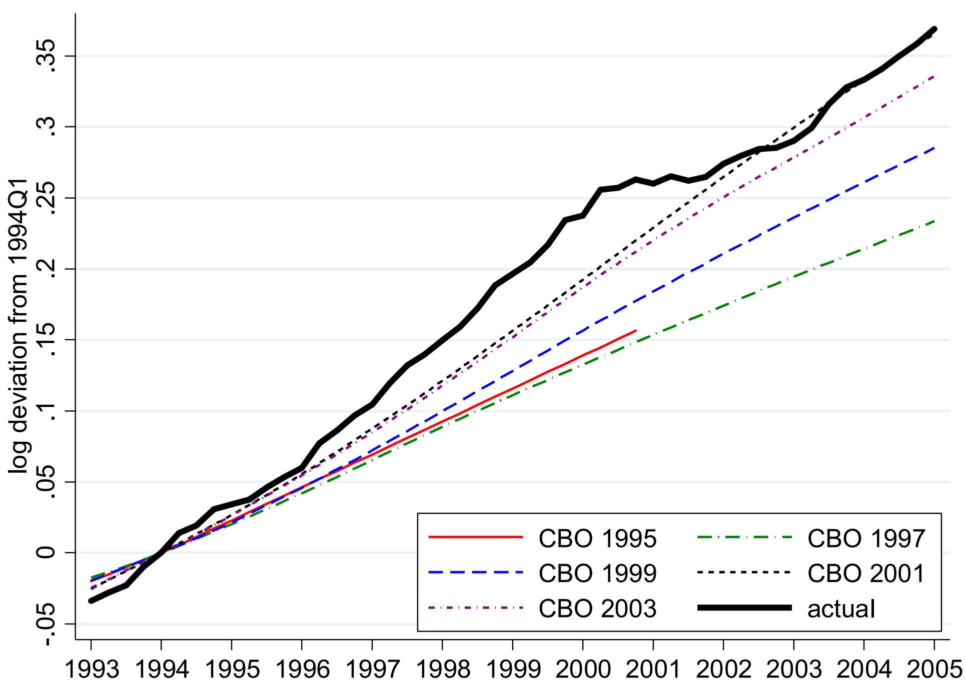
Summers, Lawrence, 2016. "The Age of Secular Stagnation: What It Is and What to Do about It," *Foreign Affairs* 95(2): 2-9.

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

Panel A: The Great Recession



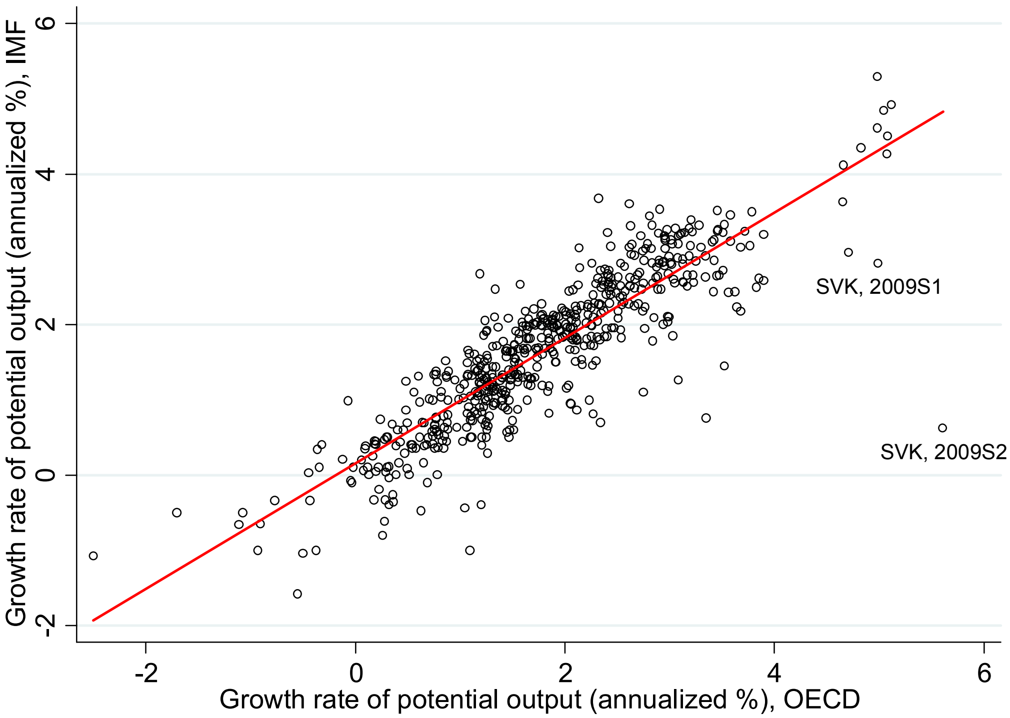
Panel B: The 1990s Productivity Boom



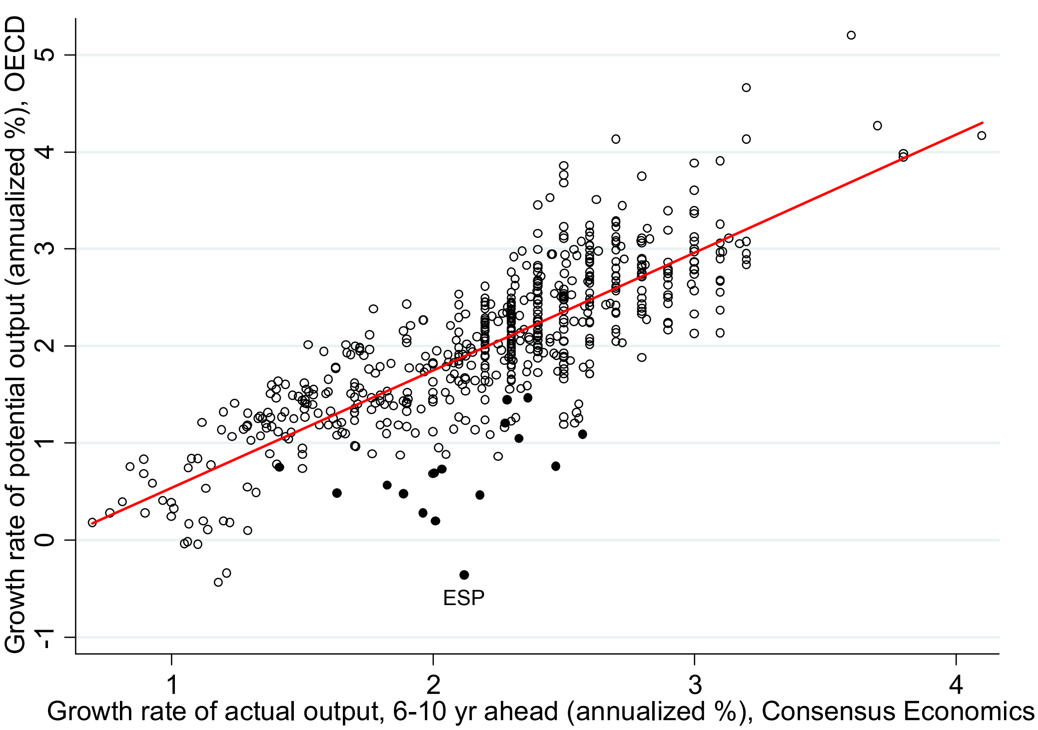
Notes: The figure plots estimates of U.S. potential output from the Congressional Budget Office made at different time periods. The solid black line represents real GDP in the U.S.

Figure 2: 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

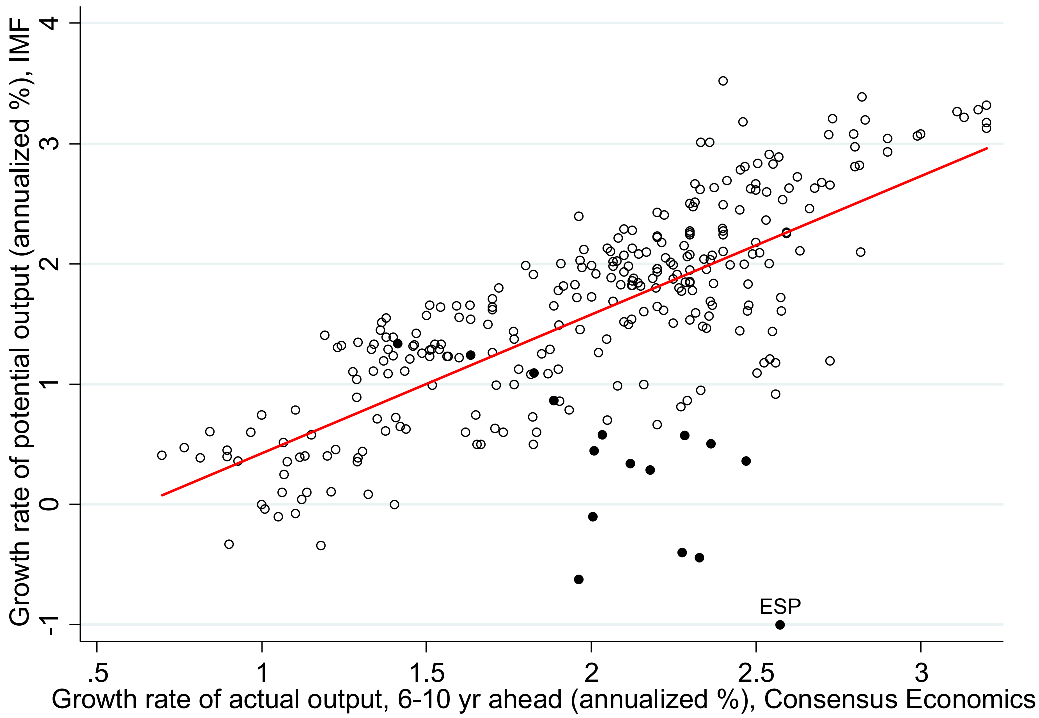


Panel B. OECD vs. Consensus Economics



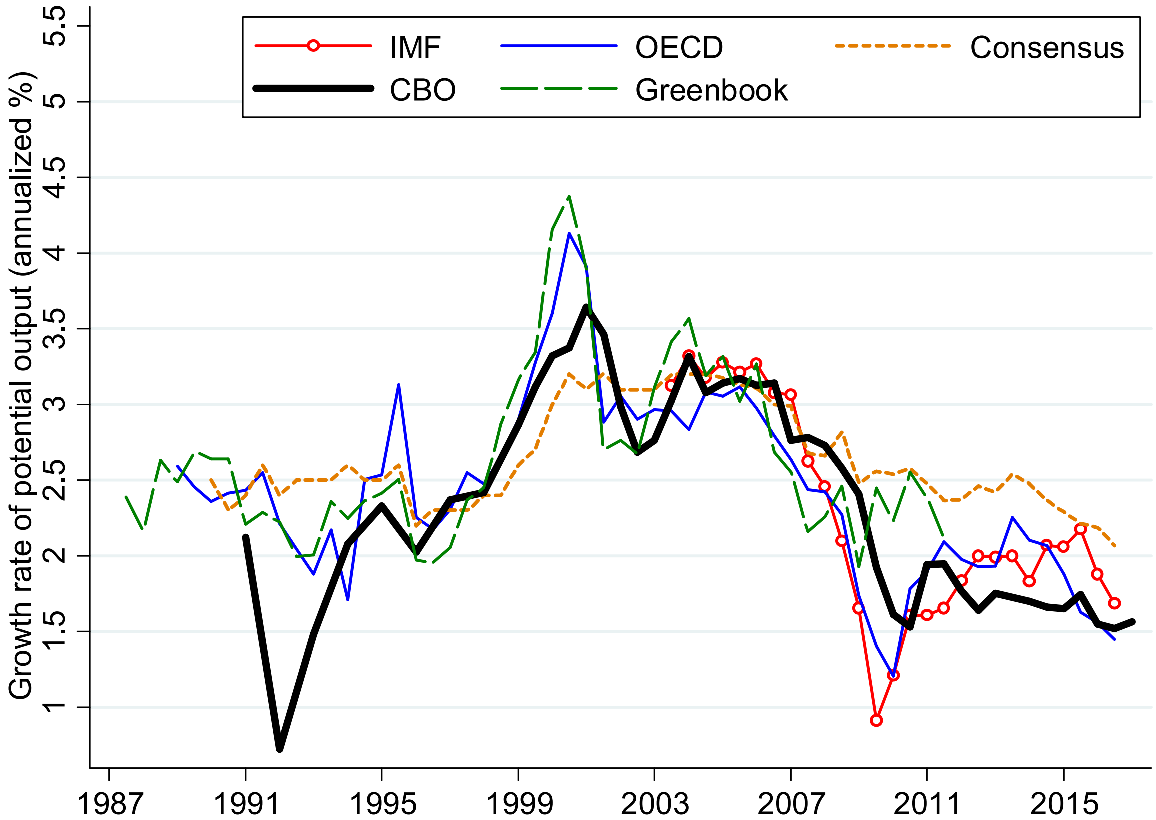
(continued on next page)

Panel C: IMF vs Consensus Economics



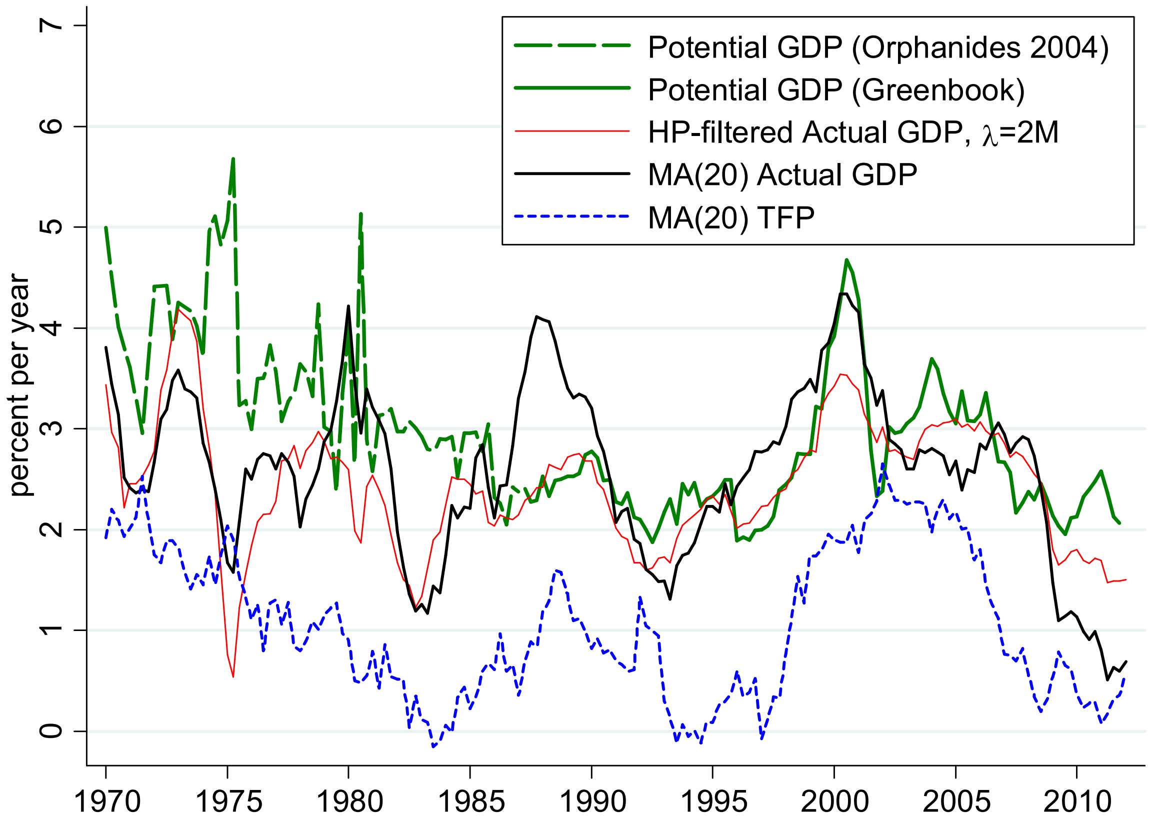
Notes: Filled markers in Panels B and C show observations for Spain in the 2009-2016 period.

Figure 3: Comparison of estimates potential output growth rate and forecasted long-term growth for actual output, USA.



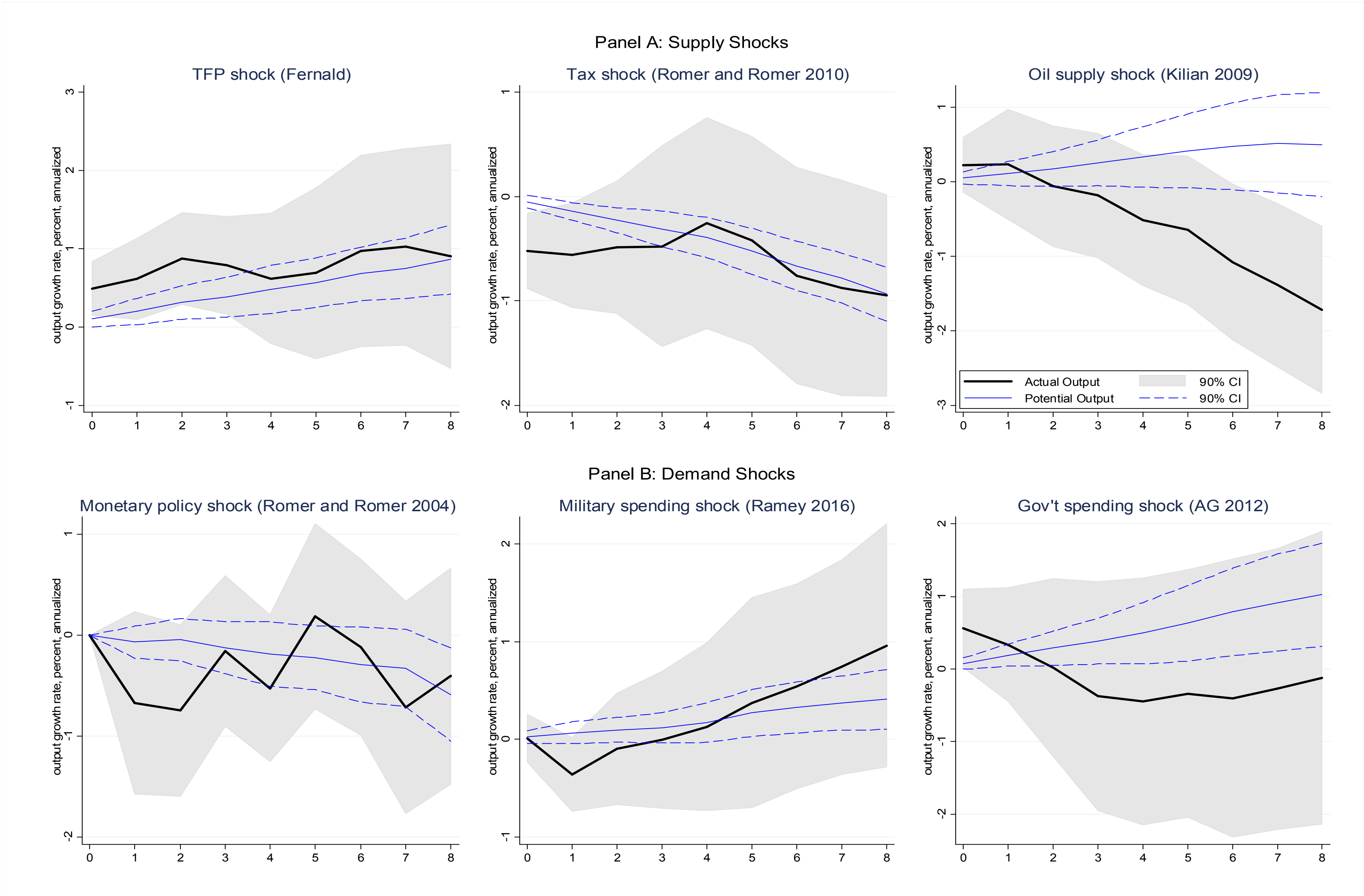
Notes: All series in the figure are based on real time data. All series are at the semi-annual frequency. The Potential output for IMF, OECD, and CBO is reported for the current calendar year. Potential output for Greenbooks is the semiannual average of quarterly growth rates of potential output for 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 4: Real-time estimates potential output growth rate and trends in actual ou tput growth rate, USA.



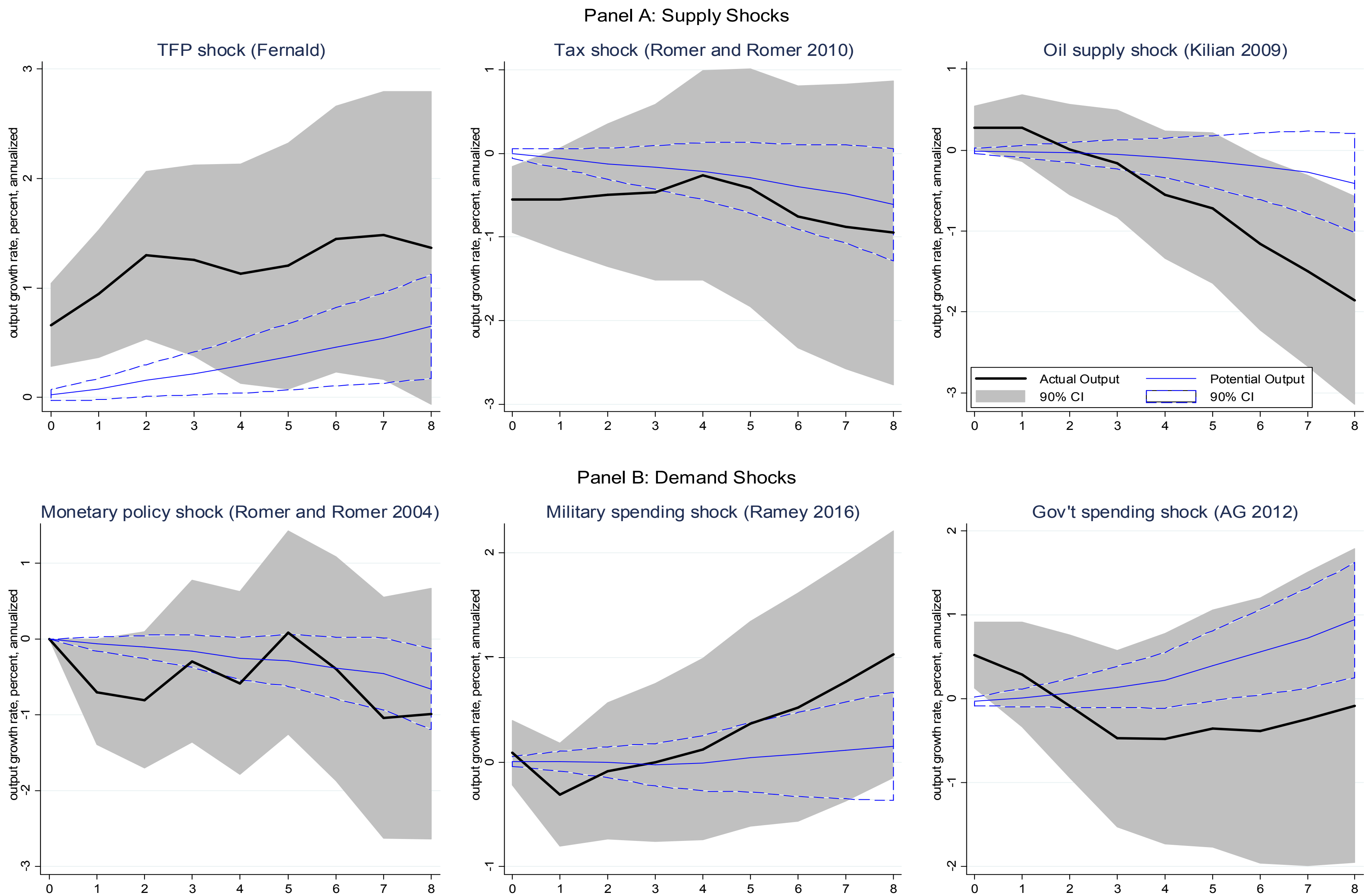
Notes: All series in the figure are based on real time data. All series are at the quarterly frequency. Potential output for the pre-1987 period is taken from Orphanides (2004). Potential output for 1987-2011 is from the Federal Reserve Bank of Philadelphia. Potential output is measured as the growth rate of potential output between a given quarter and the next 3 quarters. *HP-filtered actual output* 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 2,000,000. *MA(20) 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. *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. We use the latest vintage of TFP data.

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



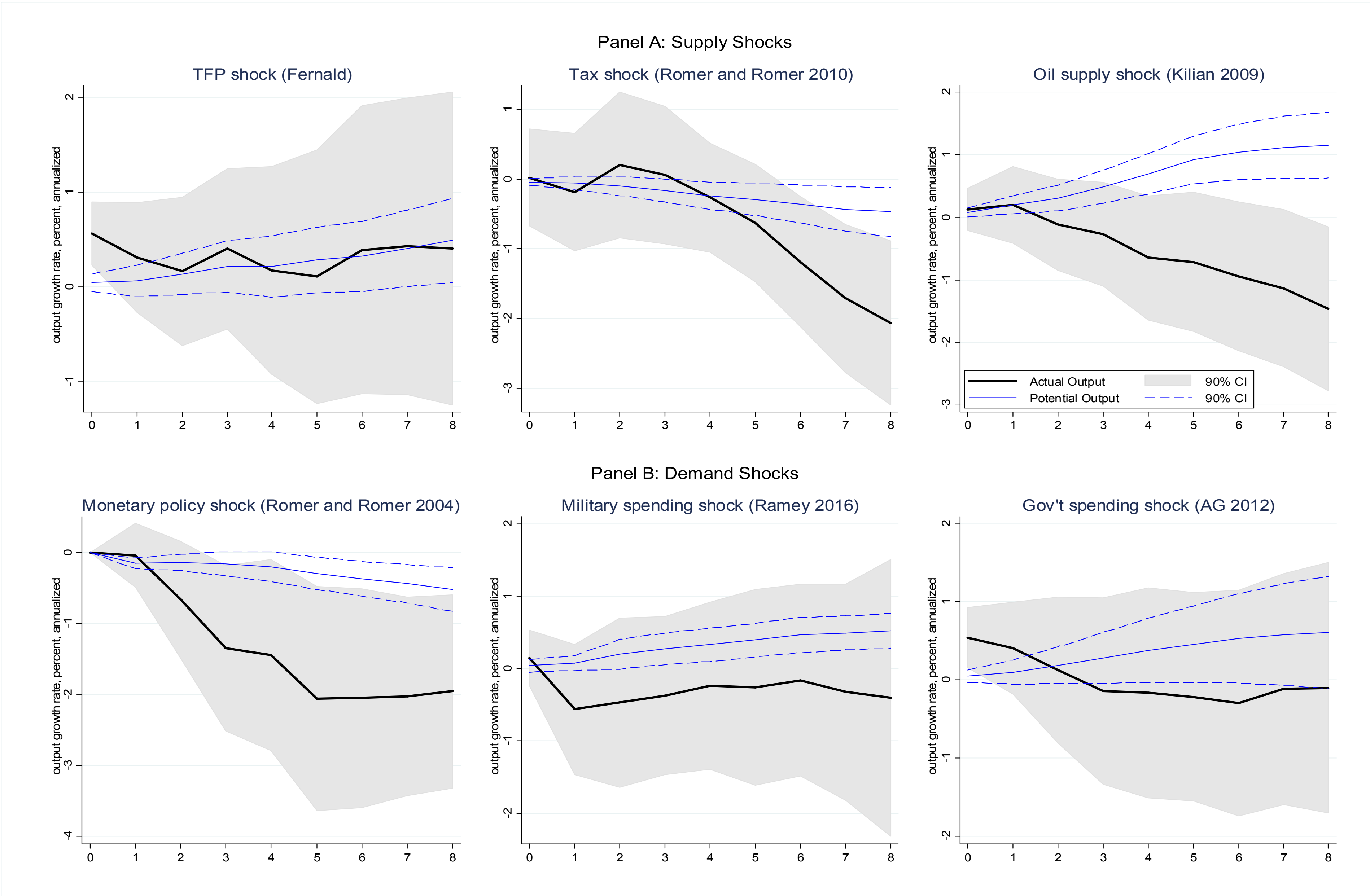
Notes: The figure reports impulse response functions (IRFs) estimated using equations (2) and (3). The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) available at the Federal Reserve Bank of Philadelphia.

***Figure 6*: Responses of Output and Greenbook Estimates of Potential Output in U.S. to Shocks: ADL specification.**



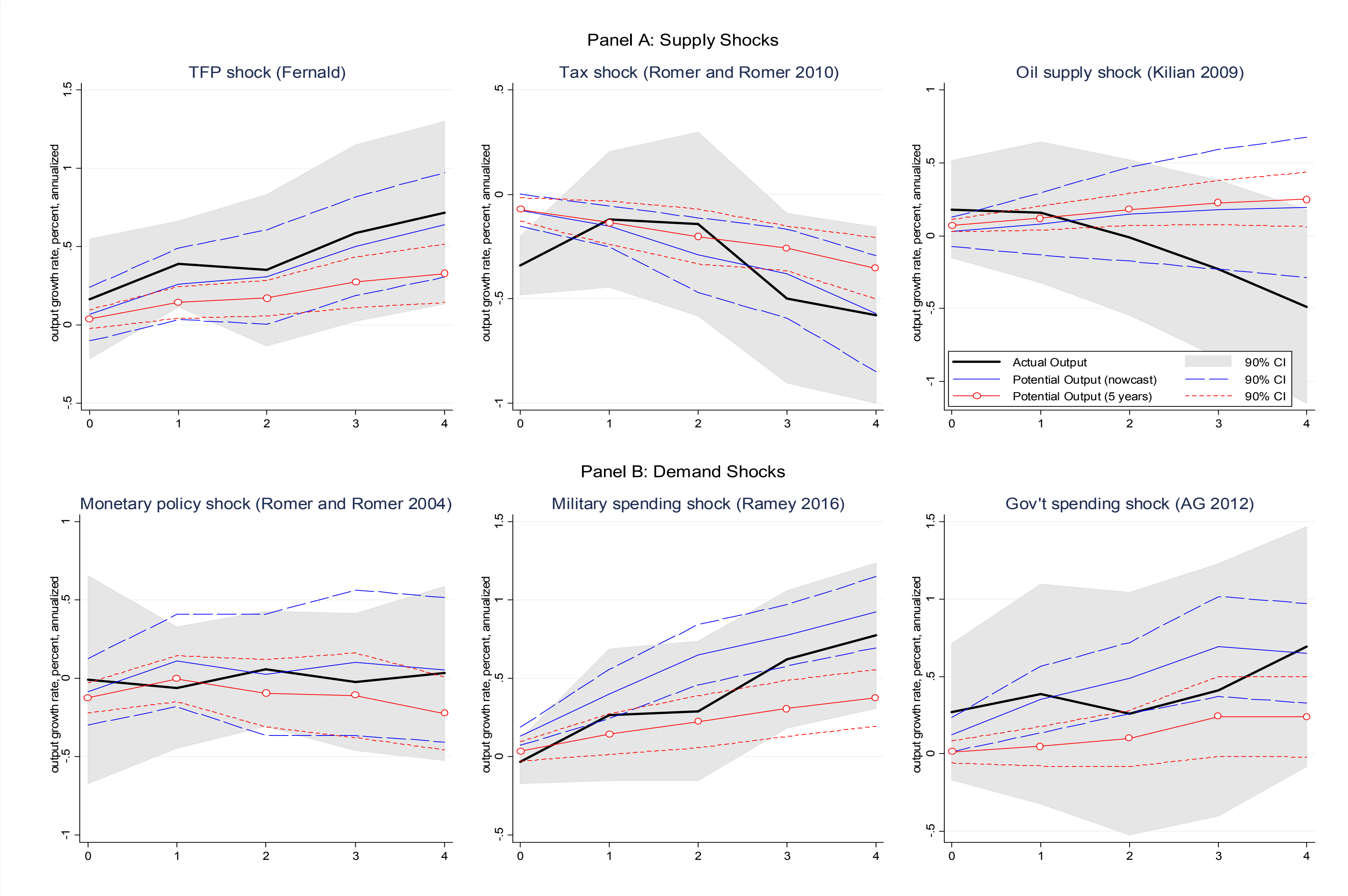
Notes: The figure reports impulse response functions (IRFs) estimated using equation (4), which is an auto-distributed lag specification. The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) available at the Federal Reserve Bank of Philadelphia.

Figure 7: Responses of Output and Greenbook Estimates of Potential Output in U.S. to Shocks: Extended Sample.



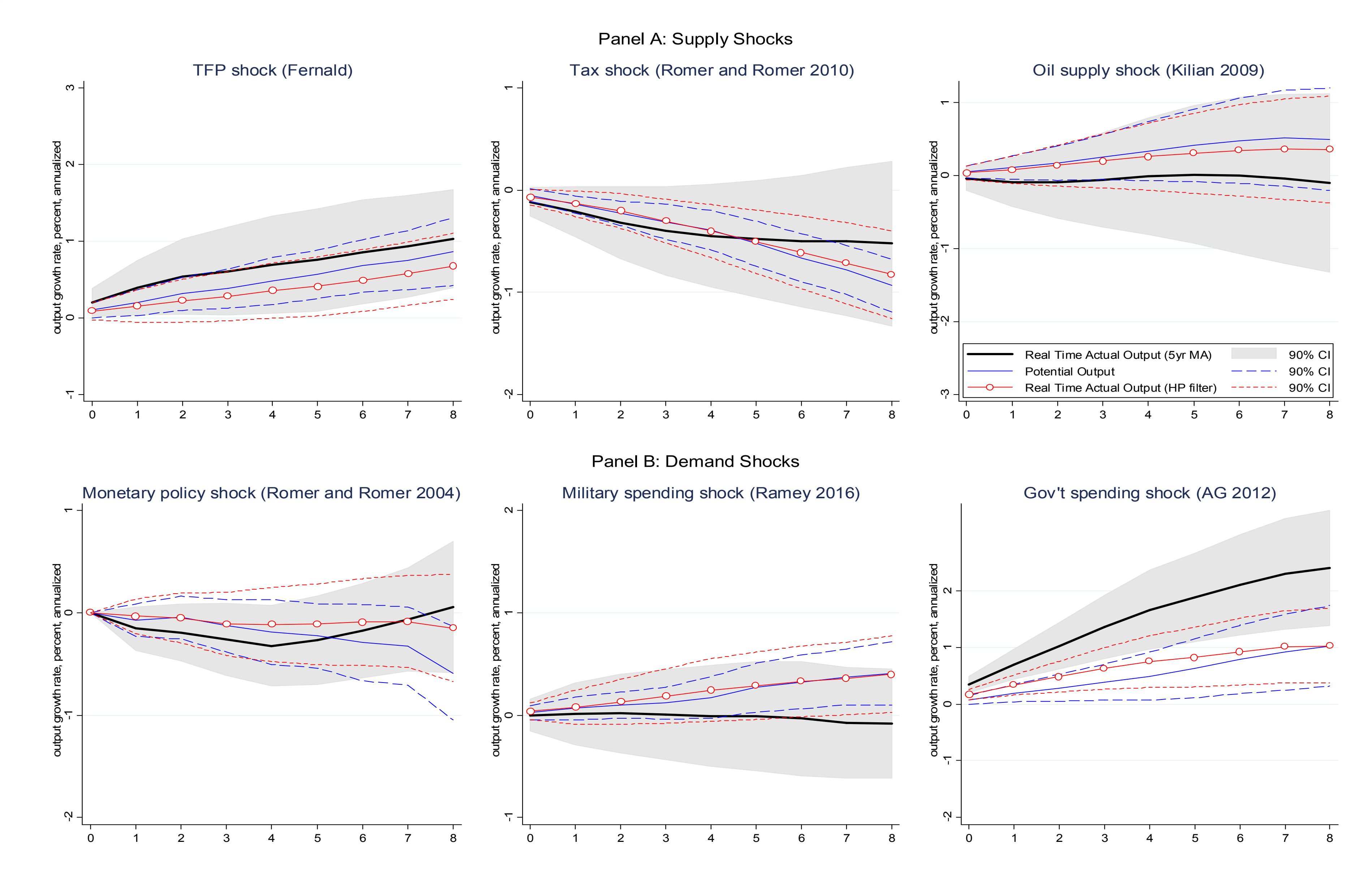
Notes: The figure reports impulse response functions (IRFs) estimated using equation (2) and (3). The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) using output gap data starting in 1970.

Figure 8: Responses of Output and CBO Estimates of Potential Output in U.S. to Shocks.



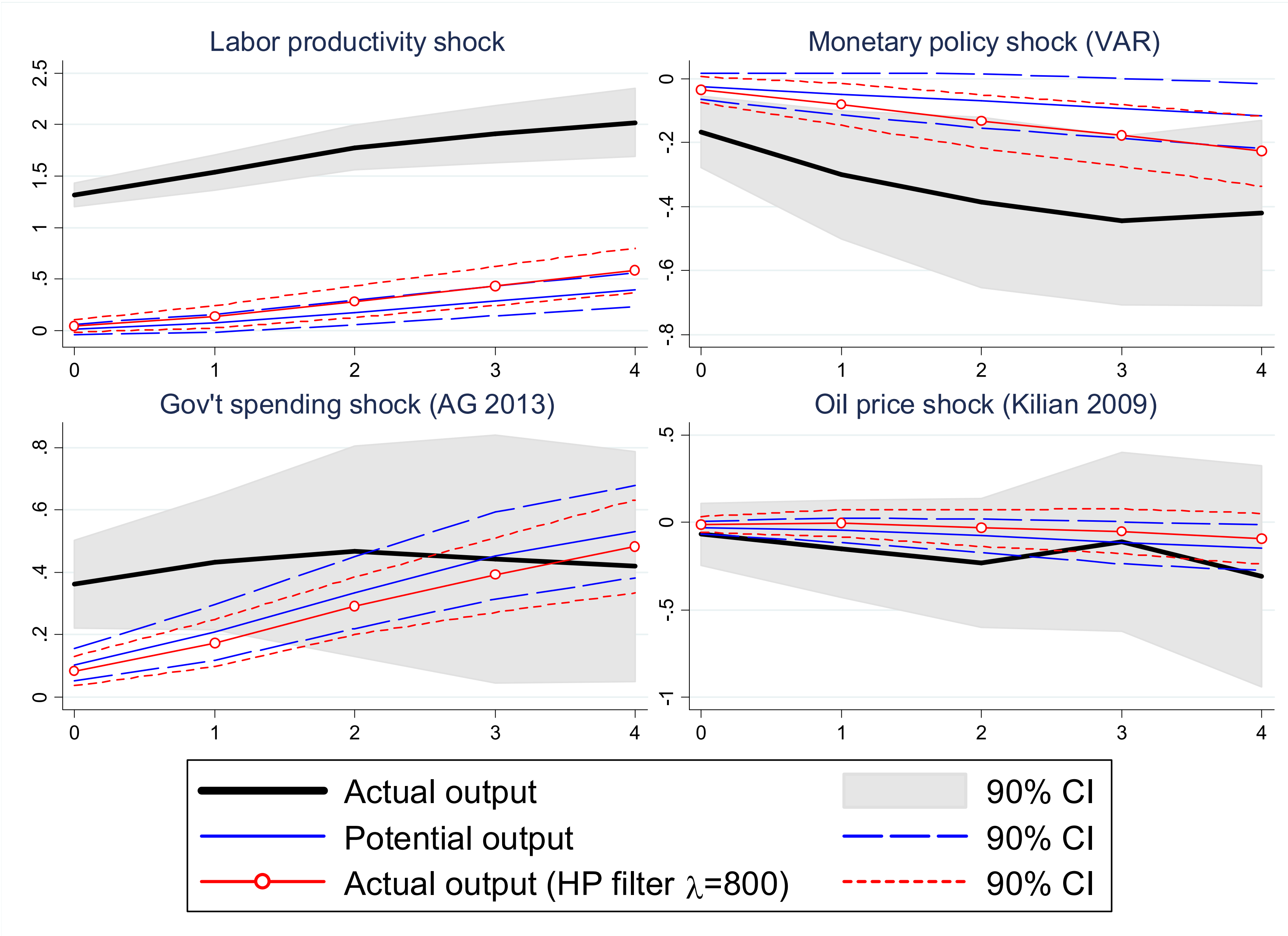
Notes: The figure reports impulse response functions (IRFs) estimated using equations (2) and (3). The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) available from the Congressional Budget Office.

***Figure 9*: Responses of Greenbook Estimates of Potential Output, Moving-Average of Output, and HP-filtered Output in U.S. to Shocks.**



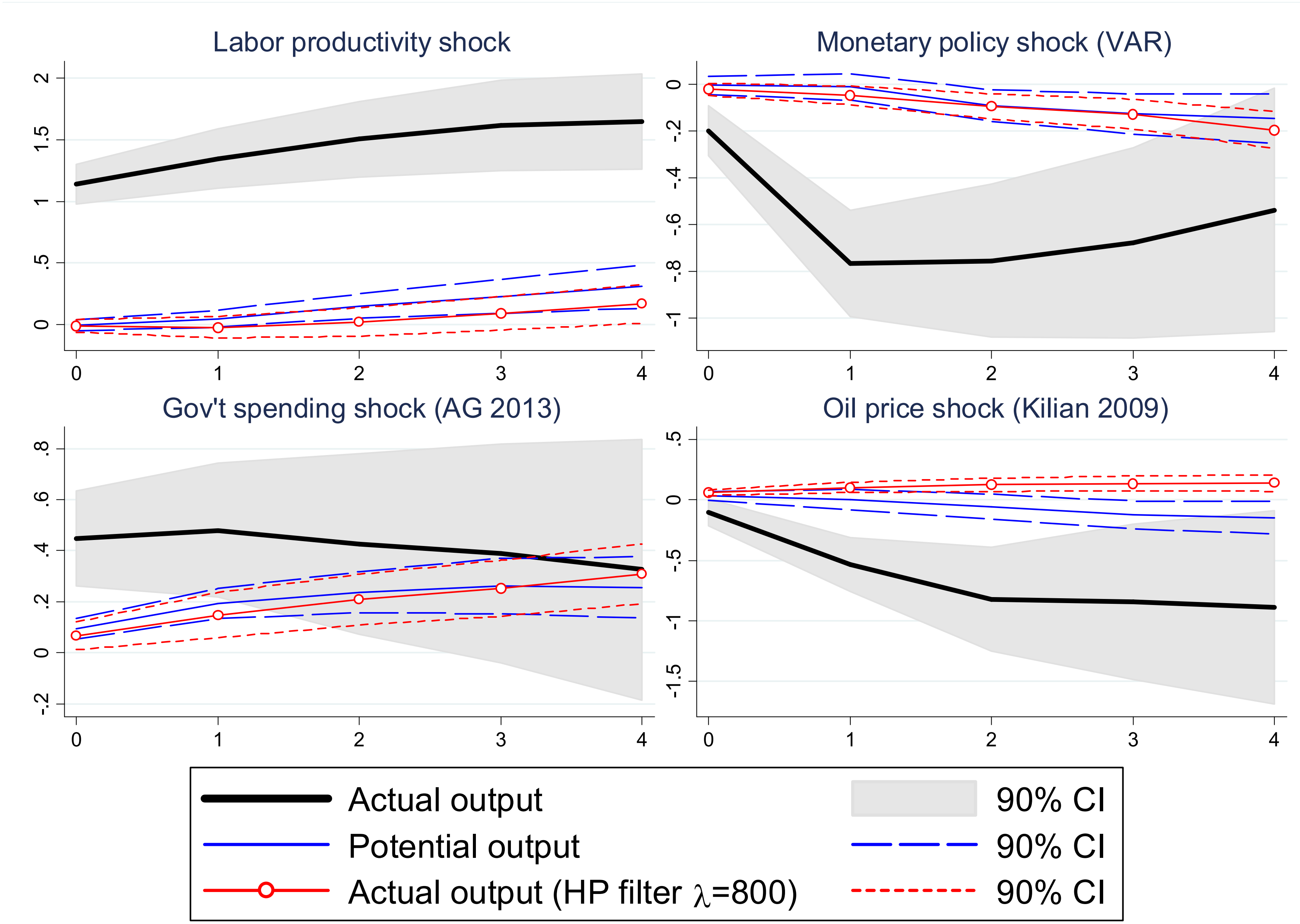
Notes: The figure reports impulse response functions (IRFs) estimated using equations (2) and (3). The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) available at the Federal Reserve Bank of Philadelphia. HP-filtered actual 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 2,000,000. 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.

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



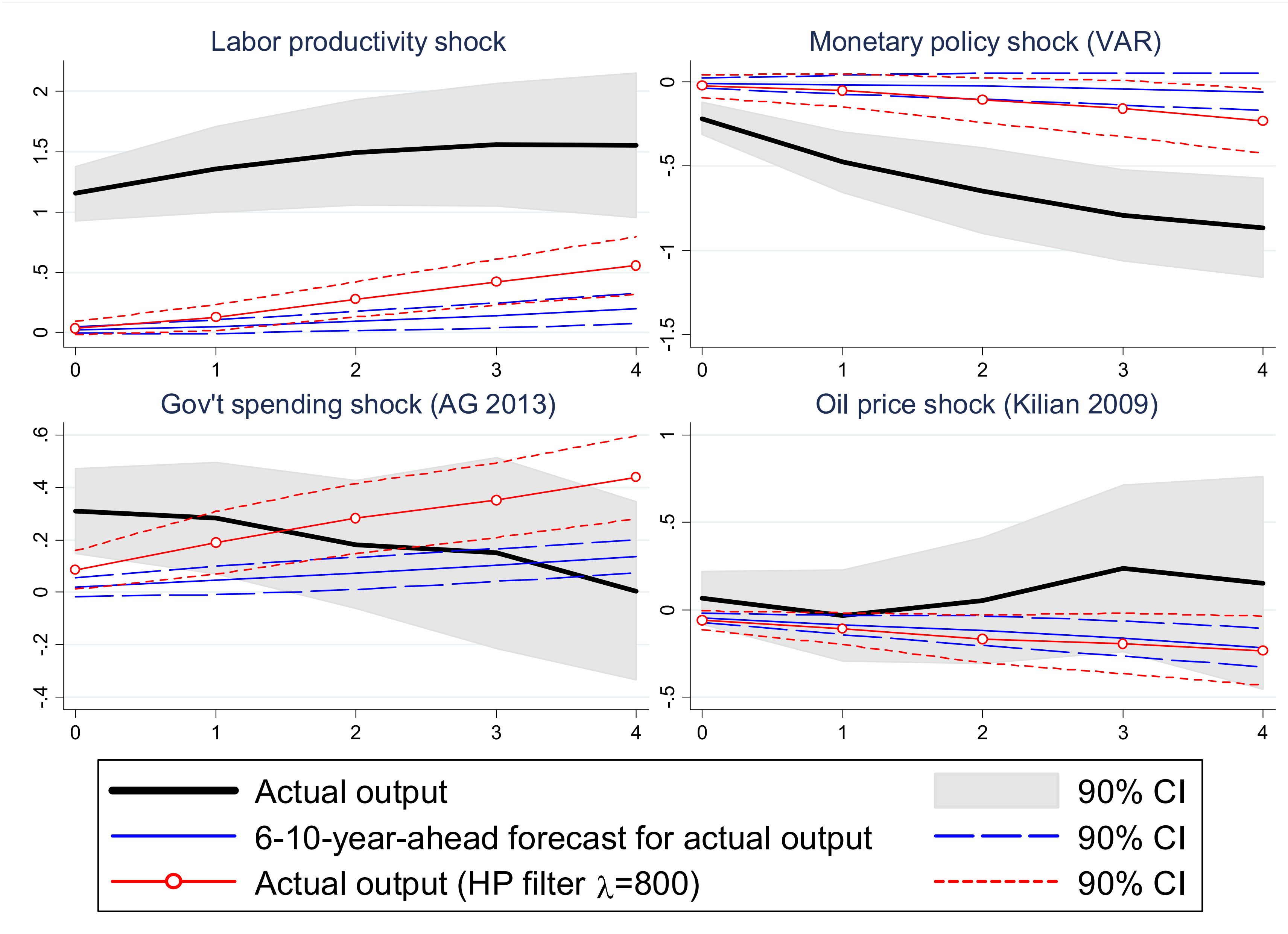
Notes: The figure shows impulse response functions (IRFs) for growth rates of actual and potential output (nowcast). IRFs are estimated using equations (5) and (6). The horizontal axis measures time in semesters (6 months). The vertical axis measures growth rate of output per year.

Figure 11. Response of the growth rate for actual output and IMF's measure of potential output (nowcast).



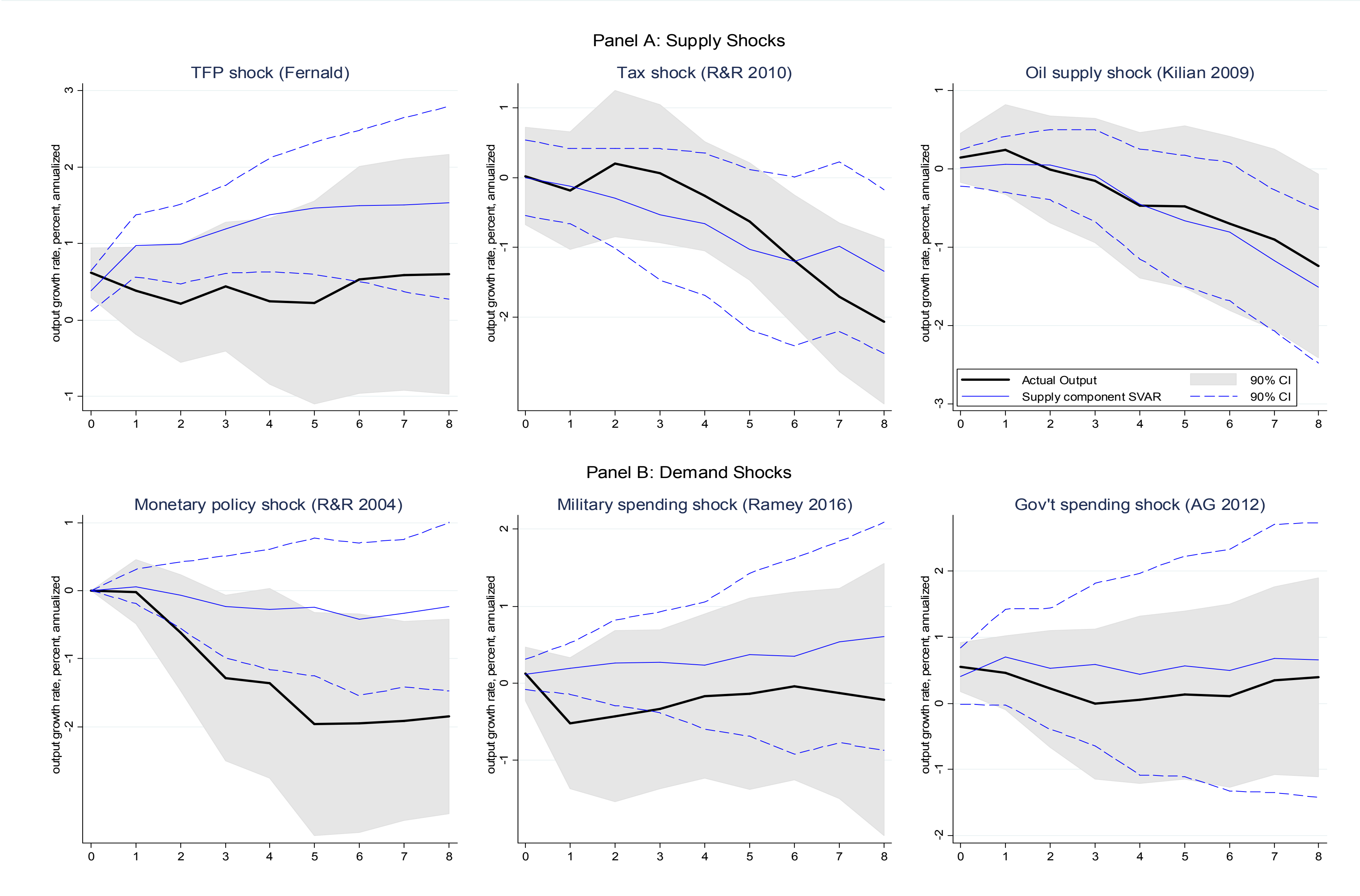
Notes: The figure shows impulse response functions (IRFs) for growth rates of actual and potential output (nowcast). IRFs are estimated using equations (5) and (6). The horizontal axis measures time in semesters (6 months). The vertical axis measures growth rate of output per year.

Figure 12. Response of the growth rate for actual output and Consensus Economics' 6-10-year ahead forecast for actual output.



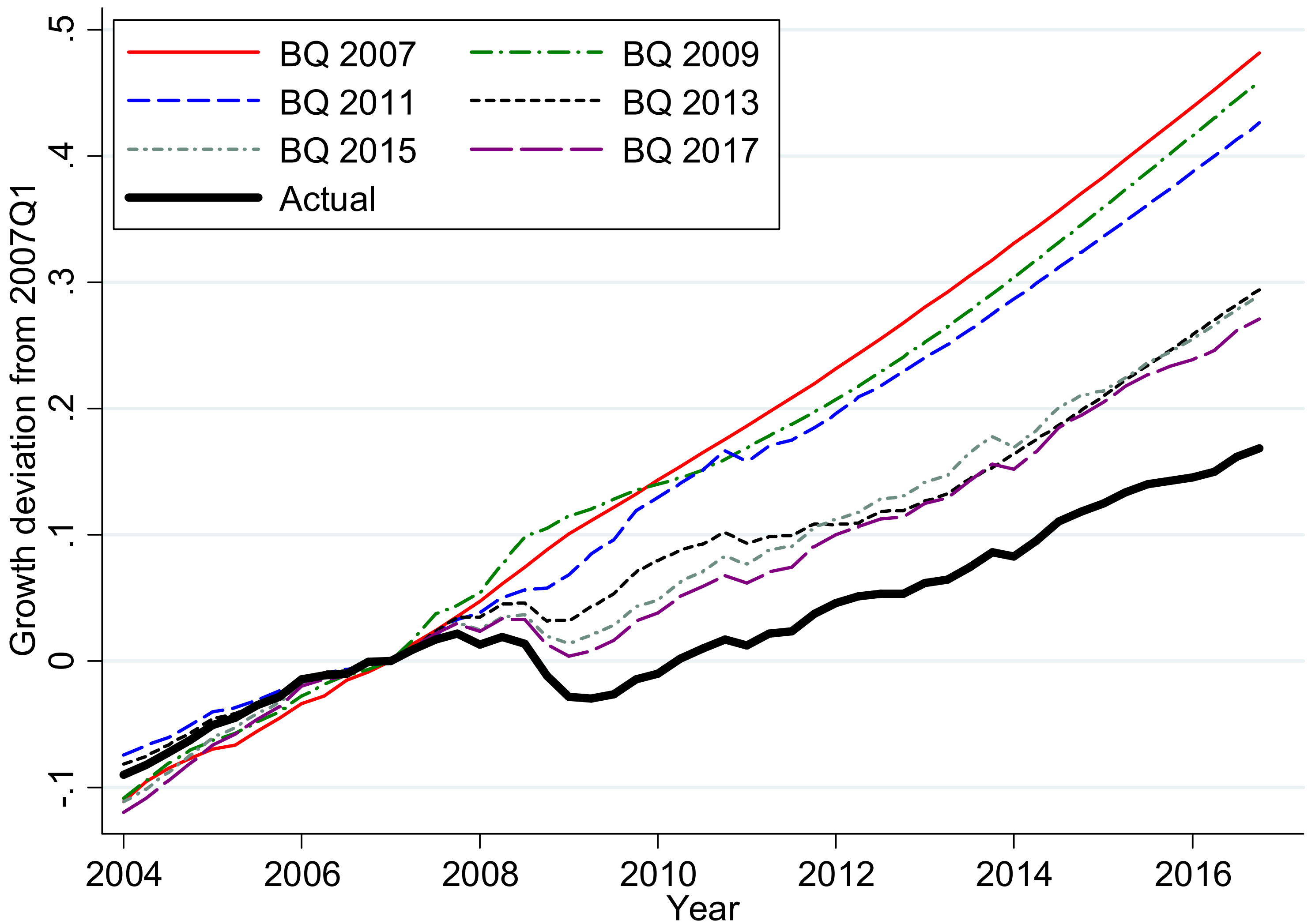
Notes: The figure shows impulse response functions (IRFs) for growth rates of actual output and 6-10-year ahead forecast for actual output growth rate (Consensus Economics). IRFs are estimated using equation (5) and (6). The horizontal axis measures time in semesters (6 months). The vertical axis measures growth rate of output per year.

Figure 13: 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). Supply component SVAR 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.

Figure 14: Revisions in Estimates of U.S. Potential Output from Blanchard-Quah (1989) approach.



Notes: The figure plots estimates of U.S. potential output made at different time periods using the Blanchard and Quah (1989) methodology estimated in real-time. The solid black line represents real GDP in the U.S.

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.

Table 2. 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 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.

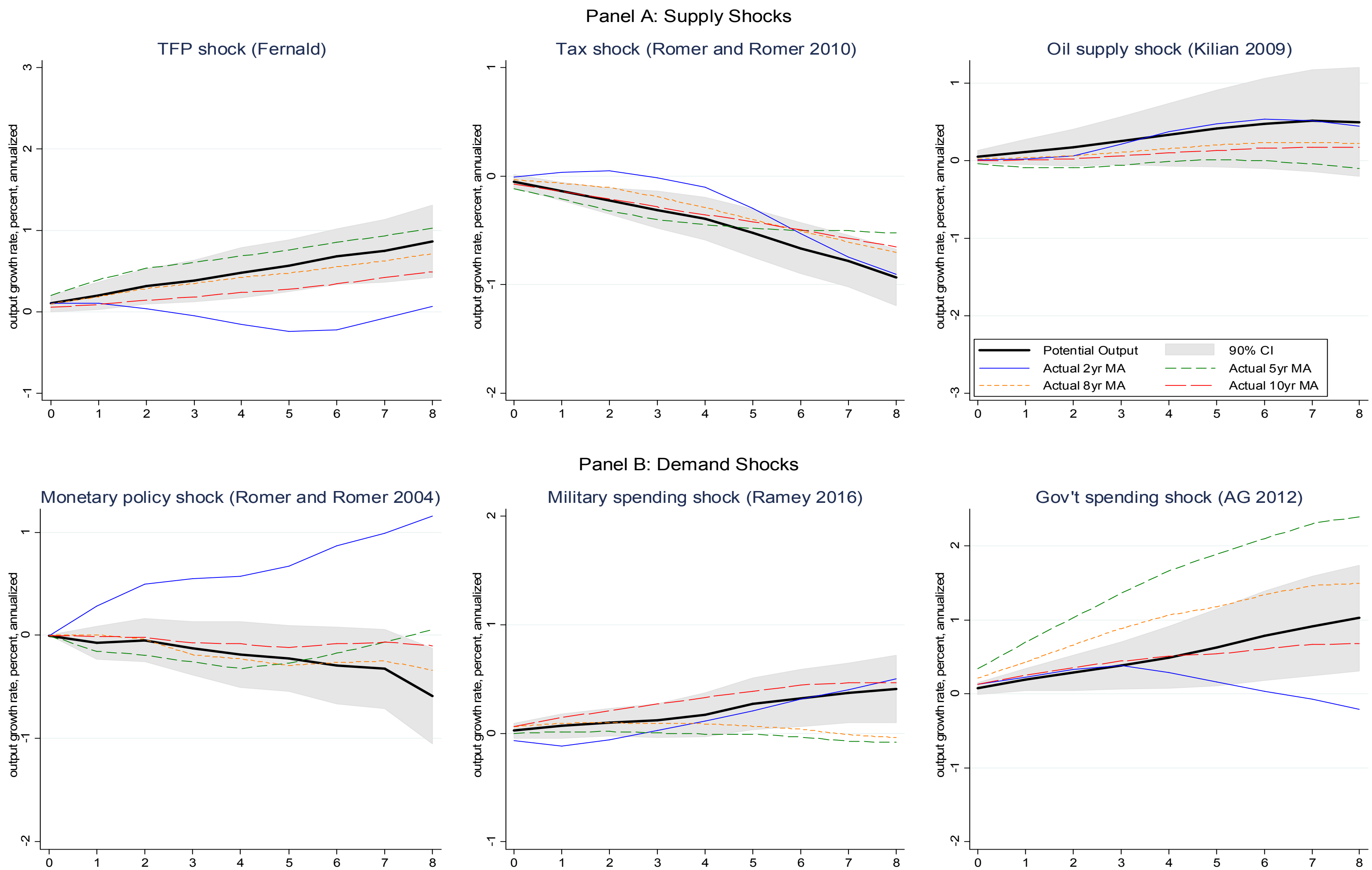
Table 3. Predictability of Revisions in Estimates of Potential GDP.

Dependent variable: $\left(\log Y^*_{t\|t} - \log Y^*_{t\|t-1}\right)$	Source						
	CBO	CBO	Greenbook	Greenbook	OECD	IMF	Consensus Economics
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\left(\log Y^*_{t-1\|t-1} - \log Y^*_{t-1\|t-2}\right)$	0.204	0.235	0.294***	0.301***	-0.066	-0.154***	-0.355***
	(0.132)	(0.144)	(0.086)	(0.085)	(0.040)	(0.044)	(0.045)
Dummy 9/11		-0.250***		-0.697***			
		(0.064)		(0.034)			
Dummy Great Recession		-0.108***		0.031			
		(0.034)		(0.092)			
Observations	42	42	96	96	1,282	548	566
R-squared	0.065	0.126	0.085	0.137			
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) and, in columns (2) and (4), dummy variables equal to one in 2001Q3 (Dummy 9/11) and 2008Q1 through 2009Q2 (Dummy Great Recession) and zero otherwise. 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 (5)-(6) are across countries and include time and country fixed effects.

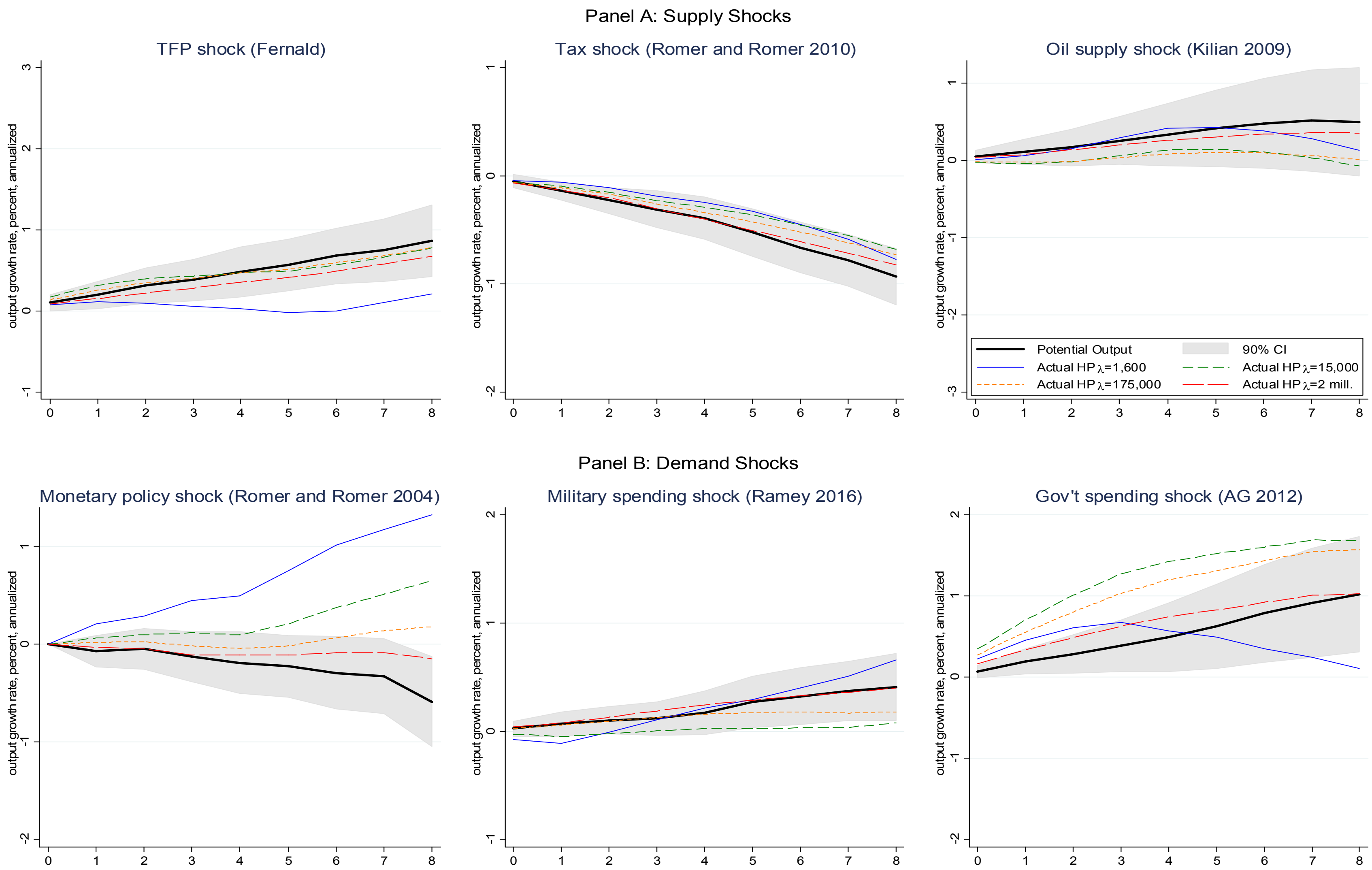
Appendix

Appendix Figure 1: Responses of Moving-Averages of Real-Time U.S. Output to Shocks.



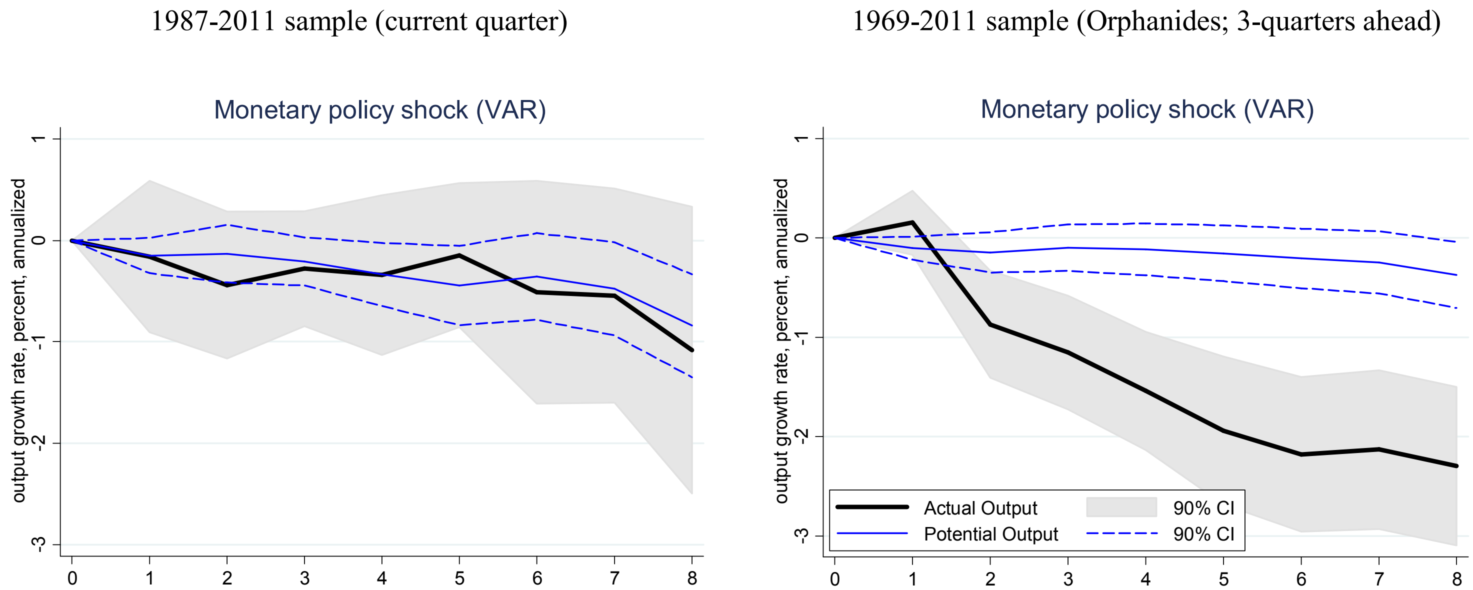
Notes: The figure reports impulse response functions (IRFs) estimated using equations (2) and (3). The estimation sample covers the longest possible period with non-missing observations for shocks and potential output (output gap) available at the Federal Reserve Bank of Philadelphia.

Appendix Figure 2: 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 3: 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 Table 1. P-values for tests for U.S. data

Shocks	Actual output		Potential output		Equality of IRFs for 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						
TFP shock	0.020	0.296	0.169	0.002	0.120	0.972
Government spending shock, (AG 2012)	0.065	0.922	0.462	0.042	0.025	0.369
Tax shock (RR 2010)	0.002	0.106	0.000	0.000	0.000	0.974
Military spending shock (Ramey 2016)	0.000	0.204	0.320	0.020	0.000	0.528
Oil price shock (Kilian 2009)	0.018	0.012	0.871	0.187	0.023	0.002
Monetary policy shock (RR 2004)	0.409	0.533	0.638	0.042	0.444	0.808
Panel B. Greenbook, 1969-2011						
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)	0.405	0.059	0.007	0.000	0.146	0.002
Monetary policy shock (RR 2004)	0.000	0.018	0.001	0.005	0.344	0.796
Panel C1. Greenbook, 1987-2011, 5yr MA						
TFP shock	0.441	0.016	0.169	0.002	0.971	0.962
Government spending shock, (AG 2012)	0.041	0.001	0.462	0.042	0.314	0.065
Tax shock (RR 2010)	0.977	0.868	0.000	0.000	0.225	0.076
Military spending shock (Ramey 2016)	0.955	0.218	0.320	0.020	0.426	0.016
Oil price shock (Kilian 2009)	0.967	0.296	0.871	0.187	0.184	0.001
Monetary policy shock (RR 2004)	0.313	0.461	0.638	0.042	0.000	0.013
Panel C2. Greenbook, 1987-2011, 5yr MA real time						
TFP shock	0.488	0.008	0.169	0.002	0.955	0.601
Government spending shock, (AG 2012)	0.004	0.000	0.462	0.042	0.056	0.011
Tax shock (RR 2010)	0.973	0.285	0.000	0.000	0.503	0.363
Military spending shock (Ramey 2016)	0.999	0.794	0.320	0.020	0.866	0.095
Oil price shock (Kilian 2009)	0.993	0.893	0.871	0.187	0.874	0.110
Monetary policy shock (RR 2004)	0.806	0.884	0.638	0.042	0.000	0.008
Panel C3. Greenbook, 1987-2011, HP real time						
TFP shock	0.514	0.010	0.169	0.002	0.904	0.269
Government spending shock, (AG 2012)	0.205	0.010	0.462	0.042	0.096	0.771
Tax shock (RR 2010)	0.089	0.001	0.000	0.000	0.759	0.570
Military spending shock (Ramey 2016)	0.779	0.078	0.320	0.020	0.452	0.804
Oil price shock (Kilian 2009)	0.998	0.419	0.871	0.187	0.357	0.326
Monetary policy shock (RR 2004)	0.998	0.640	0.638	0.042	0.000	0.001
Panel D. CBO, 1991-2011						
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 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 3 for details.

Appendix Table 2. P-values for tests for international data

Shocks	Actual output		Potential output		Equality of IRFs for 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						
TFP shock	0.000	0.000	0.042	0.010	0.000	0.000
Oil price shock (Kilian 2009)	0.003	0.089	0.046	0.241	0.021	0.136
Monetary policy shock (VAR)	0.001	0.159	0.002	0.050	0.001	0.200
Government spending shock, (AG 2012)	0.033	0.364	0.001	0.002	0.087	0.826
Panel B. IMF, HP filter, real time						
TFP shock	0.157	0.025	0.042	0.010	0.341	0.662
Oil price shock (Kilian 2009)	0.012	0.001	0.046	0.241	0.004	0.000
Monetary policy shock (VAR)	0.043	0.002	0.002	0.050	0.301	0.688
Government spending shock, (AG 2012)	0.031	0.000	0.001	0.002	0.250	0.458
Panel C. OECD						
TFP shock	0.000	0.000	0.005	0.000	0.000	0.000
Oil price shock (Kilian 2009)	0.045	0.728	0.113	0.007	0.073	0.661
Monetary policy shock (VAR)	0.007	0.003	0.148	0.011	0.103	0.053
Government spending shock, (AG 2012)	0.001	0.092	0.000	0.000	0.001	0.541
Panel D. OECD, HP filter, real time						
TFP shock	0.123	0.010	0.005	0.000	0.000	0.000
Oil price shock (Kilian 2009)	0.215	0.823	0.113	0.007	0.042	0.776
Monetary policy shock (VAR)	0.038	0.001	0.148	0.011	0.013	0.000
Government spending shock, (AG 2012)	0.001	0.000	0.000	0.000	0.755	0.876
Panel E. Consensus Economics						
TFP shock	0.000	0.000	0.249	0.022	0.000	0.000
Oil price shock (Kilian 2009)	0.117	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, HP filter, real time						
TFP shock	0.046	0.001	0.249	0.022	0.050	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.