IMF Working Paper

How Informative Are Real Time Output Gap Estimates in Europe?

by Alvar Kangur, Koralai Kirabaeva, Jean-Marc Natal, and Simon Voigts

INTERNATIONAL MONETARY FUND

IMF Working Paper

European Department

How Informative Are Real Time Output Gap Estimates in Europe?

Prepared by Alvar Kangur, Koralai Kirabaeva, Jean-Marc Natal, and Simon Voigts[1]

Authorized for distribution by Jörg Decressin and Enrica Detragiache

September 2019

Abstract

We study the properties of the IMF-WEO estimates of real-time output gaps for countries in the euro area as well as the determinants of their revisions over 1994-2017. The analysis shows that staff typically saw economies as operating below their potential. In real time, output gaps tend to have large and negative averages that are largely revised away in later vintages. Most of the mis-measurement in real time can be explained by the difficulty in predicting recessions and by overestimation of the economy's potential capacity. We also find, in line with earlier literature, that real-time output gaps are not useful for predicting inflation. In addition, countries where slack (and potential growth) is overestimated to a larger extent primary fiscal balances tend to be lower and public debt ratios are higher and increase faster than projected. Previous research suggests that national authorities' real-time output gaps suffer from a similar bias. To the extent these estimates play a role in calibrating fiscal policy, over-optimism about long-term growth could contribute to excessive deficits and debt buildup.

TABLE OF CONTENTS

I. INTRODUCTION

How informative are IMF-World Economic Outlook (IMF-WEO) output gap estimates for countries in the euro area? This is an important question since the output gap—the deviation of output from its potential—is often considered a key input to stabilization policy. The output gap (a summary measure of slack) typically informs central banks and fiscal authorities about the appropriateness of their policy stance, with larger slack calling for more stimulus. And yet, the output gap is notoriously hard to measure since potential output—commonly defined as the maximal level of output attainable without stoking inflationary pressures (Okun, 1962)—is unobservable.

Given its importance for policymaking, vast efforts have been put into estimating the output gap, with methodologies ranging from simple univariate and purely data-driven statistical filters to complex, micro-founded structural models. While some methodologies may perform better along specific dimensions, they all face limitations when it comes to estimating output gaps in real time. Real-time estimates are the relevant measures—since policy decisions about macroeconomic stabilization cannot wait—but they are always subject to considerable uncertainty and are revised *ex post* once more information becomes available.

This paper provides an analysis of the IMF's real-time output gap estimates for countries in the euro area. We observe that real-time output gap estimates drawn from the WEO database tend to have large and persistently negative means. For several large euro area countries the average real-time output gap is close to -2 percent of potential output over the period 1994–2017. The negative real-time gaps are even larger before the Great Financial Crisis (GFC) and tend to be systematically "revised away" in later WEO vintages, pointing to persistent negative bias in real time. Such patterns are qualitatively in line with experiences of real-time estimates by other institutions as well as with earlier research on revisions of real-time output gap estimates for the euro area, the U.S. and OECD countries.[2] Earlier research at the IMF suggests that the country authorities' estimates of real time output gaps suffer from similar bias (see Eyraud and others, 2018). To the extent that fiscal authorities rely on output gap estimates to assess the cyclical position of the economy, a large and persistent negative real-time bias could lead to systematic over-optimism about future output and fiscal revenues and thus to a larger-than-planned accumulation of public debt. Large and persistent negative real-time bias could also lead to unwarranted monetary stimulus. All in all, our paper cautions against an excessive focus on output gap estimates in real time when calibrating counter-cyclical policy.

But how do large and persistent negative real-time output gap estimates come about? To shed light on this question, we use the IMF multivariate filter (MVF, developed by Benes and others, 2010) to construct time series of artificial real-time output gap estimates based on all information available to staff in real time (data as well as forecasts). These artificial output gap estimates show what staff estimates would have been in real time if they had used the MVF filter in the past. This natural benchmark allows to isolate the respective roles of data

---

[2] ECB Monthly Bulletins (Feb 2005, Nov 2011), Tosetto (2008), Marcellino and Musso (2011), and Kempkes (2014) cover the euro area. Orphanides and van Norden (2002), Edge and Rudd (2016) focus on the U.S., whereas Tosetto (2008) and Turner and others (2016) on the OECD countries. For example, Kempkes (2014) documents considerable upward ex-post revisions in output gap estimates by the EC, the IMF and the OECD.

revisions, forecast uncertainty, and judgment in explaining the negative bias in real time output gap estimates.

- *Data revisions.* Following Orphanides and others (2000, 2005) we compare output gap estimates based on data available in real time against "quasi-real-time" estimates that are based on the final data vintage truncated at the year at which the gap is estimated. We corroborate the long-established result that data revisions play only a minor role in explaining large and persistent negative real time output gap estimates.

- *Forecast errors.* WEO growth forecasts tend to be too optimistic on average because staff baseline forecast is a modal forecast that typically fails to predict recessions and other tail events: a result established in earlier literature (see Box 3 for details). This has important implications for the estimation of potential output since overly optimistic forecasts tend to pull up the estimates of potential output. We estimate that based on our MVF benchmark, forecast errors lowered the real-time output gap estimates by about 0.7 percentage points on average in the euro area.

- *Judgment.* When estimating output gaps, economists typically incorporate specialized knowledge that is not directly captured by the benchmark modelling framework. To analyze the role of judgment we compare actual WEO real-time output gap estimates to MVF estimates based on WEO data and forecasts that were actually available in real time. Because we rely on a particular benchmark (MVF), what we call "judgment" may reflect staff's reliance on a different methodology in the past as well as any additional off-model information that could inform the assessment of potential output. The results reveal that judgment typically increased potential output by almost 1.0 percentage point of potential GDP in our sample, accounting for just above one-half of the overall tendency towards negative output gap estimates in real time.

Drawing from the IMF-WEO database, we show that systematic upward revisions to real-time output gaps are positively correlated with both public debt levels and public debt WEO forecast errors in the main countries of the euro area. Similarly, there is a robust, and negative, empirical association between primary fiscal balances and revisions to the WEO output gap estimates, controlling for a variety of other relevant variables.

We also establish that real time output gap estimates are generally not robust predictors of inflation, even though the Phillips curve relationship is clearly established using final estimates of the output gap (see also Abdih and others, 2018). Unlike real time output gap estimates, real time WEO inflation forecasts are not significantly revised and therefore are informative for monetary policy purposes. Real-time output gaps at best contain information on directional changes in business cycles and inflation, while final or real time inflation is the best predictor of future inflation.

The rest of the paper is structured as follows. Section 2 reviews the most common methods to estimate the output gap and their key properties. Section 3 presents key stylized facts showing the persistent and often large negative bias in WEO real time output gap estimates for the countries in the euro area, while section 4 decomposes the output gap bias, attributing most of it to judgment and forecast errors. Section 5 looks at the role of real time output gap

in predicting inflation in real time and related implications for monetary policy. Section 6 analyses the correlation between the upward revisions in real-time output gap estimates and debt buildup. Section 7 concludes and maps out possible avenues to minimize the real-time output gap bias and associated policy errors.

II. Estimation Methods and Desirable Properties

A. Estimation Methods

Different approaches to measuring the output gap largely reflect variations in the concept of potential output. From a purely statistical perspective, potential output can be seen as the trend component of actual output. In economic terms, potential output may be characterized by the level of utilization of the production factors consistent with stable inflation. The concept of "sustainable" output (IMF, 2015), on the other hand, refers to a level of GDP that the economy can sustainably produce over the medium term without stoking imbalances. Reflecting these variations, the most commonly used methods to estimate potential output and the output gap range from purely statistical univariate filters to fully structural Bayesian estimation of large Dynamic Stochastic General Equilibrium (DSGE) models. All these methods have conceptual or practical strengths and weaknesses (see Blagrave and others, 2015; Álvarez and Gómez-Loscos, 2018).

Univariate filters rely on purely statistical methods to extract a trend component from GDP time series. The most popular univariate filters are the *Hodrick and Prescott* (HP) filter (Hodrick and Prescott, 1997) and the *band pass* filters by Baxter and King (1999) and Christiano and Fitzgerald (2003). The filtered output is a smoothed GDP series that is interpreted as potential GDP. The appeal of univariate filters is in their simplicity and transparency; they do not require assumptions about the structure of the economy and instead rely on exogenous parameters to determine the degree of smoothness of a time varying trend. However, the filtered "trend" may not be consistent with the above-mentioned economic interpretation of potential output and may be of little relevance for policy purposes. Univariate filters also suffer from the well-known end-point problem that typically results in large revisions of recent years' estimates of potential output (Blagrave and others, 2015).

The *production function* estimation is a supply-side approach that decomposes output into its structural and cyclical components. The production factors include labor, capital, and total factor productivity (TFP), which is calculated as a residual. Potential output is derived by individually filtering labor supply and TFP. While this approach allows to decompose potential growth into its input contributions, thus providing an economic interpretation of potential, the estimates can suffer from the same end-point problem as univariate filtering. Moreover, as is the case with univariate filters, the production function technique ignores important information on labor market and inflation that is central to an economic definition of slack (measured as the gap between output and potential output).

Multivariate filters are typically built around economic relationships such as the Okun's Law and the Phillips curve, and may incorporate financial or other variables. The *multivariate filter (MVF)* presented in Benes and others (2010) and Blagrave and others (2015) embodies basic economic theory principles, in that it features a Phillips Curve which relates inflation to

the output gap, as well as an Okun's law linking the output gap to the unemployment gap. Potential output and NAIRU are estimated simultaneously, in line with Okun's (1962) concept of the output gap, and are therefore readily interpretable for the calibration of countercyclical policies.

Multivariate filters are very flexible and can in principle incorporate variables that are only loosely related to the concept of business cycle in the conventional sense. In this vein, Borio and others (2014) augment the multivariate filter with real credit and real house price growth—variables directly related to the housing cycle—in an attempt to capture macroeconomic stability more broadly. The resulting trend is then closer to a concept of *sustainable* level of output—attainable without stoking imbalances—than to the traditional, business cycle concept of potential output. Results (reported in IMF, 2015) show that at times of credit and housing booms, the sustainable level of output tends to be below potential output, leading to higher than conventional output gaps and suggesting that estimated potential output may not be sustained. IMF (2015) also caution that real-time identification of sustainable output is very difficult, and in practice the concept of sustainable output is best used as a "fire alarm" to complement the conventional output gap estimates.

Structural model-based approaches include estimated DSGE models that allow for joint estimation of structural shocks and potential output (see Vetlov and others, 2011). The output gap is typically closed in the steady state and consistent with stable inflation. As potential output is derived under flexible wages and prices, output can deviate from potential due to nominal rigidities. While theoretically the soundest approach, the results can be highly sensitive to specific modelling structure and parametrization.

B. Properties of Output Gap Estimates

As potential output and output gap are unobservable, their reliability and usefulness for policymaking can typically be assessed against the following statistical properties.

Stability (Size of Ex-Post Revisions)

Forecasts of observable variables are usually assessed ex-post by their *accuracy*, i.e. by their distance to actual realizations. Since output gaps are unobservable, another criterion needs to be introduced. We substitute *stability*—which we interpret as the average size and variation of ex-post revisions of real-time output gaps—for *accuracy*. The literature, documents that real-time output gap estimates tend to be too negative and substantially revised upwards over time. The ex-post revisions to real-time output gap estimates for the euro area, the U.S. and OECD countries are often at least as large as the estimates themselves and are characterized by a high degree of persistence (see Box 2). Stability is thus a key property for output gap estimates. The more stable the output gap, the smaller the policy error associated with its misestimation. For example, in the EU output gap estimates are used in structural fiscal balance rules within the stability and growth pact mechanism (SGP). Since past fiscal policy actions cannot be "revised" in hindsight, unstable and downward biased output gap estimates may result in involuntary buildups of debt and/or costly adjustments.

Efficiency

Output gap estimates are *efficient* if they utilize all information available at the time of estimation—in line with the concept of forecast efficiency (Nordhaus, 1987). If an estimate is efficient, future revisions should not be predictable and past revisions should not help explain subsequent revisions, i.e. revisions should not be serially correlated. This property also contributes to *stability* as efficient estimates should undergo smaller ex-post revisions.

Asymptotic Zero-Mean

All filter-based methods, e.g. the HP filter or the MVF, imply that the estimated output gap has zero-mean asymptotically. This is to reflect the notion that gaps are temporary and should sum to zero, consistent with an economy growing along a balanced growth path in the long term. Similarly, ex-post revisions should be distributed around a mean of zero. One-sided revisions reflect a *systemic bias* in the real-time output gap estimates. If the bias is negative, for example (one-sided positive revisions), structural fiscal deficits may be underestimated in real time, potentially leading to higher than desired debt buildup, which would increase the risk of procyclical adjustment in the future. Several studies document systematic negative biases in real-time output gap estimates, with some noting potential negative consequences in terms of fiscal sustainability (see Section III.C).

The zero mean output gap property is a general condition traditionally associated with the steady-state of the economy (e.g. in the Phillips curve) and is consistent with traditional business cycle analysis in which cycles are broadly symmetric. It is also embedded in output gap estimates that feed into standard policy reaction functions, highlighting that the long run zero mean is a desirable property in practice as it serves as an anchor in particular for fiscal policies.[3] A long-run zero mean is also reflected in the empirical assessments of output gaps by international institutions (see Section III.B). It is only under a zero mean that the average structural fiscal balance (i.e. the fiscal balance adjusted for the output gap) is in line with the average fiscal position over the cycle: if the cycle is symmetric, the structural balance gives the average fiscal position over the cycle, and fiscal rules can be anchored around the structural balance.

In contrast, if output gap estimates are on average negative, fiscal rules based on structural balances would result in excessive deficits. Indeed, separate strands of business cycle literature have recently questioned the symmetry of the business cycle, pointing out that there may be good reasons why countries may occasionally spend more time below potential. This could occur, for example, in case of downward nominal wage rigidities leading to asymmetric business cycle responses that have to be assessed in conjunction with potential effects of hysteresis (see Box 1). Also, stabilization policies can suffer from asymmetric limitations—e.g. ZLB on monetary policy (that going forward can be magnified by low neutral real interest rates, see Rachel and Summers, 2019) or constraints on fiscal policies imposed by high debt—which makes managing overheating easier than providing stimulus during recessions.

---

[3] To the extent monetary policy seeks to stabilize output, output gap properties serve as an anchor also for monetary policy. This is less relevant in the euro area where the primary objective of common monetary policy is to maintain price stability. Country-specific stabilization objectives are often served by fiscal policies.

Economic Consistency

As real-time output gap estimates are intended to accurately signal inflationary pressures in the economy, they should be correlated with inflation and consistent with other indicators of slack such as unemployment, capacity utilization or labor market tightness. These additional indicators of slack can be incorporated directly into the MVF or used separately to assess real-time measures of the output gap.

Box 1. Nominal Wage Rigidities, Hysteresis, and Output Gap Estimates

Some DSGE models that feature asymmetric nominal wage rigidities can produce negative long-run output gaps. The output gap in those models is derived as the difference between output under nominal rigidity and output under flexible wages and prices. In the framework of Benigno and Ricci (2011) since downward wage rigidity dominates the forward-looking reaction of wage setters, the output gap is always nonpositive in the long run (it is zero without uncertainty).

Aiyar and Voigts (2019) argue that downward nominal wage rigidities will lead to a negative average output gap by inducing negative demand shocks to have larger impact on unemployment and output than positive shocks. Therefore, in their view conventional methods to estimate potential output such as HP or MVF exhibit an intrinsic *upward* bias in output gap estimates that is especially large in deep demand-driven recessions. Bound by the asymptotic zero mean property, instead of allowing for asymmetry in shock adjustment and thus in output gap estimates, these filters spuriously lower the estimate of potential.

Coibion and others (2017) find that real-time potential GDP estimates across several institutions tend to be sensitive to demand shocks and under-respond to supply shocks. Following Blanchard and Quah (BQ, 1989) they estimate a bivariate VAR with an identifying restriction that only supply shocks can have permanent effects on output. Deriving the real-time potential GDP as a historical contribution of supply shocks to growth they find that in the aftermath of the GFC the US potential output tended to be *under-estimated* and consequently the output gaps should currently be larger and more negative than estimated by most institutions.

It is debated whether demand shocks only have transitory effects on GDP. Blanchard and others (2015) find that 83 percent of recessions associated with supply shocks result in sustained decline in output. They also find that recessions triggered by demand shocks are frequently followed by lower output or even lower output growth and can thus have permanent effects. Of all the recessions associated with intentional disinflation—the purest cases of demand shocks—almost two-thirds are associated with lower long-term output. In earlier work, Blanchard and Summers (1986) and Ball (1999) questioned the assumption of long-run neutrality of money and argued that monetary policy and other demand factors can have permanent effects.

These findings suggest important hysteresis effects, implying that potential output can decline more at times of recessions and thus output gaps can be higher (or less negative). Alichi and others (2019) apply the MVF with labor market hysteresis on US data, finding significantly higher estimates of NAIRU, lower estimates of potential and substantially higher (less negative) output gaps throughout the GFC as well as 1980s compared to model simulations without hysteresis. Univariate regression by Hamilton (2018) can capture persistence and centers cycles around zero mean.

A separate line of research has shown how demand shocks can have long-run effects on aggregate output by inducing shifts in a long-run aggregate supply curve. Bashar (2011) shows that in all G-7 countries aggregate demand shocks positively affect aggregate supply shocks (through correlation of demand shocks with labor productivity ruled out in standard BQ identification) causing permanent effects on the output level.

III. OUTPUT GAP ESTIMATES THROUGH WEO VINTAGES

A. Definitions

In this paper we define the real-time output gap as the output gap estimated for year t in the Fall WEO of the same year. For example, the real-time output gap estimate for country i ($y_{i,t|t}$) for the year 2000 is taken from the 2000 Fall WEO vintage, the real-time gap for the year 2001 is taken from the 2001 Fall WEO vintage, and so forth. Therefore, real time output gap estimates are defined at a time when about half of the actual data for any particular year is available. Using this definition, we compile an unbalanced panel for euro area countries starting in 1994 when systematic reporting of the output gap estimates became available in WEO. Our data extends to the 2017 fall WEO vintage that we refer to as providing the "final" WEO output gap estimates.

Our primary question of interest is the information content of output gap estimates when policy decisions are actually made. Following Orphanides and van Norden (2002) we therefore work with contemporaneous real time gaps, estimated at time t conditional on information available at that time as these are the relevant measures for fiscal and monetary policy decisions inevitably taken in real time.[4] This also allows us to isolate different sources of real-time output gap bias, including the effects of data revisions and forecast errors.

To evaluate the quality of macroeconomic forecasts, analysts typically rely on the statistical properties of forecast errors. Since output gaps are unobservable and subject to continuous revisions in subsequent years, we examine both output gap estimates (*levels*) and revisions of output gap estimates (*changes*) against a wide set of properties, as outlined in Section II. Revisions are defined as deviations of real-time output gap from final (2017 WEO) estimates, or from WEO estimates up to five years later ($y_{i,t|t+5}$). We interpret systematic asymmetry in revisions of real-time estimates (*changes*) as "bias". Since final estimates broadly satisfy the zero-mean property, the revision bias translates into a *level* bias for real-time estimates.[5] We further investigate whether real-time estimates are consistent with other indicators of slack such as inflation and estimates of unemployment gaps. Results are presented for all euro area countries as well as for a baseline sample of 11 countries for which full data for 1994–2017 is available. Finally, we compare the statistical properties of the WEO output gap estimates for the euro area with estimates for other advanced economies, and with estimates by other international organizations (see Box 2).

---

[4] Practices in defining real-time output gaps differ depending on the questions of interest. Grigoli and others (2015) focus on ex post output gap estimates ($y_{i,t|t+1}$) in order to isolate the impact of future data revisions and uncertainty over potential output as well as ignore policy reactions and short-term forecast errors.

[5] A bias in real-time estimates (i.e. a tendency of revisions to be positive) does not require a zero mean in the final estimates as a necessary condition.

B.   Properties of WEO Real-Time Output Gap Estimates

Large and Asymmetric Ex-Post Revisions

The revisions of real-time output gap estimates tend to be large and mostly upawards. Figure 1 presents the WEO real-time output gap estimates for the period between 1994–2017 as well as the "final" ($y_{i,t\,|\,2017}$) and "5-year later" ($y_{i,t\,|\,t+5}$) revised WEO estimates.[6] Real-time output gaps tend to be revised substantially over time. For the euro area, real-time estimates are revised upwards by more than 1 percent of potential output on average, with substantial variation across countries and time periods. Throughout the WEO history, downward revisions (compared to the final 2017 estimates) constitute only about one-sixth of all observations and are attributable to very few countries. In other words, real-time estimates do not perform well against both the stability and zero-mean properties.

Figure 1. Real-Time Output Gaps Through WEO Vintages

Sources: IMF *WEO* and staff calculations.

---
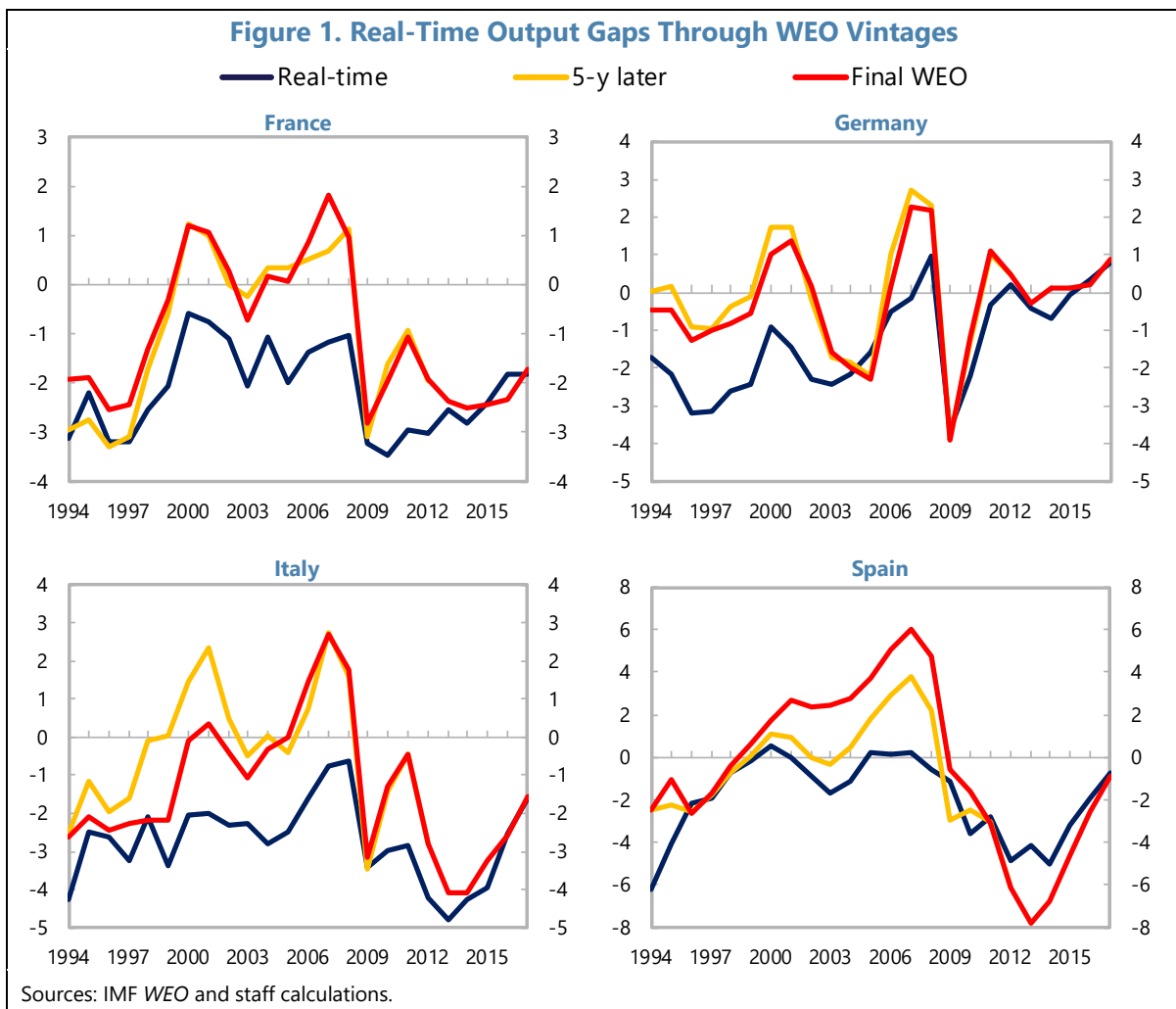
[6] As the time between real-time and final gaps becomes smaller until in 2017 real-time output gap is identical to final WEO 2017 output gap, the last few years in the sample show small revisions. Therefore, in calculating revisions we drop the last five years or compare against output gap estimates five years later with similar effect.

The real-time estimates are revised gradually over time. Table 1 presents mean and median estimates for the short- and medium-term revisions, as well as the number of countries that have one-sided (upward) revisions on average, statistically significant bias, and serially correlated revisions (see also Appendix Figures 1 and 2). While short-term revisions tend to vary with a cycle (upward in expansions and downward in recessions) for some countries, the 5-year later revisions as well as the

Table 1. Descriptive Statistics of Revisions to WEO Real-Time Output Gaps (EA11)				
	1 year	2 year	5 year	WEO'17
Mean bias	0.40	0.68	1.27	0.80
Median bias	0.40	0.63	1.32	1.08
Upward revision	11/11	11/11	11/11	9/11
(share of countries)	(100%)	(100%)	(100%)	(82%)
Statistical significance	8/11	8/11	9/11	8/11
(share of countries)	(73%)	(73%)	(82%)	(73%)
Serial correlation	1/11	6/11	8/11	10/11
(share of countries)	(9%)	(55%)	(73%)	(91%)

Sources: IMF *WEO* and staff calculations.

revisions in the 2017 vintage are predominantly upward throughout the sample period. Short-term revisions (one or two years after) are small, likely driven by data and short-term forecast revisions and do not suffer from serial correlation. Already after two years, revisions start to exhibit serial correlation that increase in subsequent years, implying that output gap estimates have not been efficient: past revisions could be used to improve future output gap estimates.

Negative Bias in Real-Time Estimates

Systematic upward revisions manifest themselves in large and persistently negative output gap estimates in real time. Indeed, over the whole WEO history, France and Italy have not recorded a single positive output gap estimate in real time, Germany has recorded four, Spain five and the euro area only one. Table 2 compares the means of output gaps estimated in real-time and in the final (2017) WEO vintage for our baseline sample of 11 countries for which longer time-series are available as well as for all euro area countries and tests whether the mean output gaps are statistically different from zero. Only in four countries—Luxembourg, Ireland, Malta, and Cyprus—are we unable to reject the hypotheses that the mean is zero. In all other countries the output gaps are statistically negative on average (p-value<0.1) and for France, Italy, Finland, Portugal and Spain are close to, or exceed, -2 percent of potential GDP. Such patterns indicate that a negative bias has historically been a persistent feature of real time output gap estimates.

The negative average output gap estimates are largely revised away over time. By the final (2017) WEO vintage, for most countries the output gaps, averaged over a comparable 1994–2017 period, are substantially more positive and not statistically different from zero (notwithstanding GFC and irregular business cycle length). We conclude that the real-time estimates are biased as is reflected in predominantly upward revisions, indicating that staff's assessment of potential tends to be more optimistic in real-time than in hindsight. Still, France and Italy as well as the weighted euro area retain large and statistically significant negative gap estimates also in the final WEO.[7]

---

[7] As unobservables, output gaps are re-estimated once new information becomes available; in this regard in 2018 Fall WEO the mean estimate of the output gap for euro area since 2010 (comparable to Table 2) is higher at -0.4 and statistically not significantly different from zero.

Table 2. Real-Time Output Gaps: Mean Estimates and Significance (1994–2017)								
	Real-Time (n=368)			Latest Vintage (n=368)			Latest Vintage (n=463)		
	Mean	SE	p-value	Mean	SE	p-value	Mean	SE	p-value
Euro Area	-1.7	0.23	0.00	-0.5	0.42	0.28	-0.6	0.32	0.08
Austria	-1.1	0.17	0.00	0.0	0.24	0.92	0.0	0.24	0.92
Belgium	-1.0	0.17	0.00	-0.3	0.20	0.20	-0.3	0.20	0.20
France	-2.1	0.18	0.00	-1.0	0.30	0.00	-1.0	0.30	0.00
Germany	-1.3	0.27	0.00	-0.2	0.29	0.39	-0.2	0.29	0.39
Italy	-2.7	0.22	0.00	-1.4	0.37	0.00	-1.4	0.37	0.00
Netherlands	-1.0	0.38	0.01	-0.3	0.36	0.47	-0.3	0.36	0.47
Finland	-2.0	0.38	0.00	0.1	0.52	0.87	-0.1	0.53	0.84
Greece	-1.5	0.85	0.09	-0.3	1.17	0.79	-0.4	1.12	0.71
Ireland	-0.3	0.58	0.60	-0.2	0.71	0.77	-0.2	0.71	0.77
Portugal	-2.2	0.33	0.00	-0.7	0.46	0.11	-0.7	0.46	0.11
Spain	-1.9	0.39	0.00	-0.4	0.77	0.61	-0.4	0.77	0.61
Luxembourg	0.6	1.12	0.58	-0.5	0.52	0.38	-0.4	0.48	0.42
Malta	0.0	0.34	0.94	0.2	0.47	0.60	0.1	0.28	0.59
Cyprus	0.2	1.61	0.90	-0.1	0.96	0.94	-0.2	0.70	0.73
Slovakia	-1.7	0.44	0.00	-1.9	0.78	0.02	0.5	0.59	0.43
Estonia	-1.9	0.80	0.02	-2.3	1.34	0.08	0.2	1.20	0.89
Latvia	-0.4	0.11	0.00	-0.5	0.10	0.00	-1.8	1.26	0.15
Lithuania	-1.0	0.36	0.01	-1.2	0.47	0.01	0.7	1.03	0.52
Slovenia	-1.8	0.77	0.02	-0.4	1.29	0.78	-0.3	0.62	0.61

Source: IMF *WEO* and staff calculations.
Note: Latest vintage estimates are provided for common sample with real-time estimates (second column) as well as for full sample (third column). The 11 baseline countries are shown in the upper part of the table. Real time output gap estimates for the euro area start from 2010, for individual countries based on data availability. For Slovakia 2018 Spring WEO is used as the latest vintage for consitency in revisions.

For newer euro area countries (e.g. Baltic states, Slovenia, Slovakia, Malta) real time output gap estimates are available only for a shorter period, much of which spans the GFC and not covering a full business cycle. The sample period for these countries is thus not sufficiently long and representative to assess the asymptotic zero-mean property. Still, in the latest WEO, which spans over fuller cycles and longer period, their output gap estimates are not statistically significantly different from zero (although the standard errors tend be large reflecting more volatile business cycles).

Similarly, the sample period 1994–2017 spans from the beginning of a recession to the end of a recession, thus covering incomplete cycles and contributing to marginally negative (although statistically not different from zero) averages in the 2017 vintage. The mean estimates over full lengths of business cycles are closer to zero (less negative) also for baseline countries. Over a longer history and full business cycles, the WEO, OECD and EC final vintage weighted output gap estimates for the euro area on aggregate are smaller, averaging -0.3, -0.3, and 0 percent of potential GDP, respectively (text chart).[8] This is in line with the asymptotic zero mean property with broadly similar variation across countries evident for all three institutions.

---

[8] Unweighted averages of the WEO, OECD and EC final vintage output gap estimates corresponding to the text chart are -0.2, -0.5, and -0.2, respectively. Calculations for the OECD cover 17 euro area countries.

Final Output Gap Estimates for the Euro Area
(Percent of potential GDP; weighted average)

— IMF (WEO) — EC (AMECO) — OECD

Sources: European Commission; IMF *WEO* and OECD.
Note: unbalanced weighted averages for euro area countries based on data availability. Starting point reflects full cycles.

Revisions to WEO Real-Time Output Gaps

Sources: IMF *WEO* and staff calculations.
Note: average WEO revisions for a baseline sample of 11 countries.

A question that often arises is whether the negative bias in real-time estimates is uniquely driven by the global financial crisis (GFC). The answer is no. Real-time output gaps have been persistently negative and revised upwards throughout the 1994–2004 period preceding the pre-GFC boom-bust cycle. As the text chart illustrates, the negative bias in real-time output gaps is not only an artifact of the crisis years. The bias is however larger in the immediate pre-crisis period (2005–2008) reflecting the difficulty in forecasting recessions and turning points in business cycles in real time. Following the GFC, growth in the pre-crisis years was widely assessed to have been fundamentally unsustainable (see Section VI).

Economic Consistency Property

The bias in the real-time output gap estimates can also be seen from other indicators of slack.

Okun's Law

The time-tested Okun's law stipulates a consistent relationship between the output gap and unemployment rate deviations from its long-run trend. A positive output gap should correspond to the unemployment rate below its long-term trend and *vice versa*. However, real-time WEO output gap estimates are largely nonpositive even at times of below-trend real-time unemployment when one would have expected actual output to exceed potential.

We look at two measures of trend: the long-run average unemployment rate and the NAIRU. In the first case, France, Italy, Finland and Portugal still exhibited real-time WEO output gaps of around -2 percent of potential GDP when unemployment was below its long-term average—prima facie, this points to an inconsistency in the gap estimate. Using the NAIRU, which accounts for possible downward revisions of structural unemployment, such an inconsistency is less visible, indicating that the bias mostly reflects overly high estimates of potential output. Nevertheless, at times when the unemployment rate is below NAIRU, real-time output gap estimates are rarely positive (see Appendix Table 1).

Figure 2. Inflation and Real-Time Output Gaps

Sources: IMF *WEO* and staff estimates.
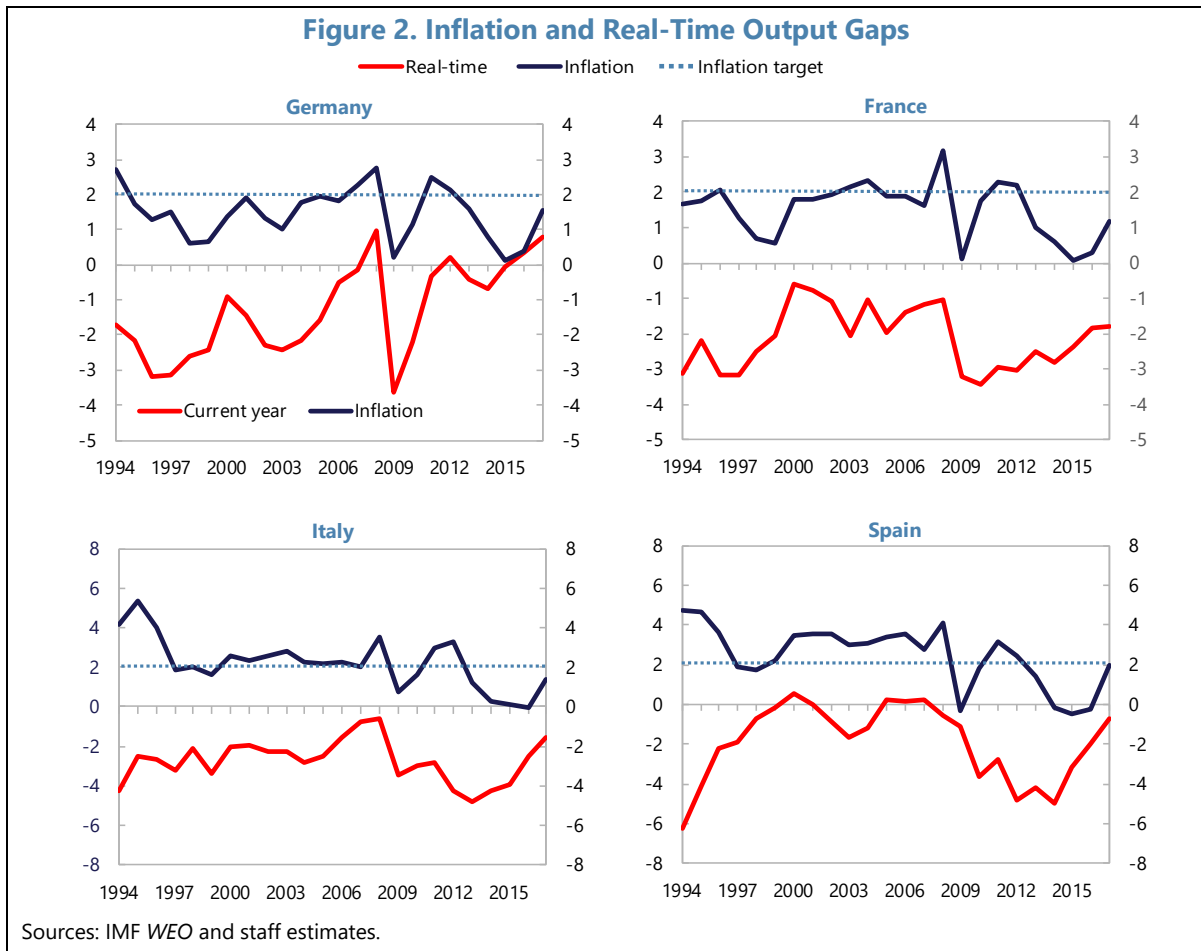
Phillips Curve

The second building block in understanding business cycles is the short run trade-off between inflation and a measure of slack such as the output gap. At times of negative output gaps inflation should fall below its expected or targeted level. However, we find a disconnect between the levels of real-time output gap and inflation. Despite persistently large negative real-time output gaps headline inflation has been hovering around or just below 2 percent in many countries (Figure 2), where also inflation expectations in the euro area have been anchored (see Abdih and others, 2018).

This disconnect is mirrored in revisions to real time inflation and real time output gaps. In our baseline sample of 11 larger euro area countries for the years 1994–2012, revisions to the Fall WEO real time output gaps have been large and systematically *upwards* (1.3 percent of potential GDP unweighted average), while revisions to real-time inflation have been marginal (-0.04 percentage points unweighted average) and, on average, downwards.[9] Therefore, if the final vintage versions of output gap estimates are useful for forecasting inflation, which we show in section V, then the real time estimates are unlikely to be, or at best can be informative in terms of the direction and not the level of inflation.

---

[9] Revisions to real time inflation forecasts in the Spring WEO vintages, when less information on monthly inflation realizations is available, average to 0.1 percentage points for the baseline sample.

C. Real-Time Estimates for Other Advanced Economies

Similar patterns can be seen in other advanced countries as well as in real-time output gap estimates by other organizations. Table 3 presents WEO mean output gap estimates for the 4 largest non-euro area countries: US, UK, Canada and Japan. For all these countries the WEO real-time output gap estimates have been significantly negative, whereas—and with the exception of Japan—the mean estimates in the 2017 Fall WEO center around zero. From 1994 to 2017 Japan has not recorded a single positive output gap estimate in real time, Canada has recorded two and the UK four. Box 2 documents similar findings of negative real-time output gap estimates and upward revisions at other international organizations.

Table 3. Real-Time WEO Output Gaps in Major Non-Euro Area Economies: Mean Estimates and Significance (1994–2017)

	Real-Time (n=96)				Latest Vintage (n=96)		
	Mean	SE	p-value		Mean	SE	p-value
US	-1.18	0.46	0.01	US	0.11	0.44	0.80
UK	-1.32	0.31	0.00	UK	-0.31	0.34	0.37
Canada	-1.27	0.26	0.00	Canada	0.17	0.24	0.49
Japan	-3.07	0.46	0.00	Japan	-1.57	0.39	0.00

Sources: IMF *WEO* and staff estimates.

Box 2. Real-Time Output Gap Assessments in International Organizations

The negative bias in real-time output gap estimates by international organizations for the euro area countries has been widely documented.

ECB (2005, 2011) and Marcellino and Musso (2011) find that real-time output gap estimates for the euro area (EC 1999–2004, OECD 2002–2010, IMF 1999–2010) are characterized by a high degree of instability, with large and predominantly positive revisions. They conclude that the information content of real-time output gap estimates tends to be low and suggest that the real-time assessment of the degree of slack in the economy should be based on a wide set of indicators. Rünstler (2002) shows that real time estimates of output gap in the euro area could be improved to a considerable extent if the cyclical co-movement between the output gap (by the ECB over 1970:Q1–2000:Q4), factor inputs and capacity utilization is exploited.

Kempkes (2014) analyze real-time output gaps for EU 15 countries over the 1996–2011 period, as estimated by the EU, the IMF and the OECD. He similarly finds a negative bias in real-time estimates (i) irrespective of the source of the data; (ii) in all real-time vintages; and (iii) across the entire cross-section of countries; the bias is estimated on average 0.5 percentage points of potential GDP per year. He notes that a systematic downward bias in real-time output gap estimates implies that structural balances tend to be over-estimated in real time and suggests that fiscal rules should incorporate ex-post checks of the unbiasedness of the cyclical components.

Hernández de Cos and others (2016) based on the EC estimates for the EU 15 over 2004–14 find that the direction of revisions to the real-time estimates (defined as an estimate of the output gap in year *t* made in Spring of year *t-1*) depends on the state of the economy: upward for expansions and downward during recessions. Based on this finding, they argue that this asymmetry should be taken into account in the computation of structural balances. Ademmer and others (2019) similarly find that for EU 28 over 2004–17 period the real-time output gap estimates are revised upwards in boom periods and downwards in recessions.

Using the OECD estimates, Tosetto (2008) and Turner and others (2016) find the revisions of real-time estimates to be large and persistent for 15 OECD countries over 2003–08 and for G7 for 2007–09. Turner and others (2016) show that additional cyclical adjustments using manufacturing capacity utilization, a share of investment in GDP as well as house prices and credit help to improve the reliability of OECD output gap estimates. Edge and Rudd (2016) and Champagne and others (2018) find that the revisions of output gap estimates have become smaller in more recent samples, based on the data on U.S. and Canada, respectively.

IV. EXPLAINING THE NEGATIVE BIAS IN REAL-TIME OUTPUT GAPS

A. Decomposing the Real-Time Output Gap Bias

This section seeks to explain what drives the large negative means of real-time output gap estimates and their sizable positive revisions. In their seminal work Orphanides and van Norden (2002) were mostly concerned with the impact of data and model or parameter uncertainty on potential output estimates in real time. They found that model or parameter uncertainty in the form of unreliable end-of-sample estimates resulted in considerable mismeasurement of output gaps in real time. Building on their work we isolate the impact of three measurable sources of bias in real-time output gap estimates: (i) data revisions, (ii) systematic bias in 5-year ahead forecasts, typically used to extend the data sample for computing real time output gaps, and (iii) judgment, i.e. deviations from the MVF estimates (capturing model or parameter uncertainty, "expert" knowledge etc.). To do so we rely on the IMF's multivariate filter (MVF, see section II.A, Benes and others, 2010 and Blagrave and others, 2015 for details) to obtain three counterfactual series of real-time output gap estimates. This allows to test each of the competing explanations in isolation by using the same methodology to process data in real time. Our sample is limited by the availability of WEO real time vintage data (GDP, inflation and unemployment). It spans the period 1990 to 2017 and covers 11 countries for which a full set of real-time data is available: Austria, Belgium, France, Finland, Germany, Greece, Italy, Ireland, the Netherlands, Portugal and Spain.[10]

Using the MVF is appealing since it has desirable stability properties, asymptotic mean-zero and a straightforward economic interpretation that also satisfies the economic consistency properties. Nevertheless, any counterfactual exercise of the sort requires benchmarks and to a certain extent is model-dependent. For example, the impact of data revisions is conditional on using the MVF and does not necessarily need to coincide with its impact when a different estimation method is used. This caveat is especially relevant for the analysis of "judgment" that is calculated by comparing counterfactual simulations against the actual WEO real-time output gap estimates and captures both discretionary expert-judgment and methodological differences between the MVF and other filtering methods used at that time.[11]

B. Data Revisions

The impact of data revisions on output gap estimates is relatively small. Following Orphanides and Norden (2002), we analyze it by comparing quasi-real-time estimates of the output gap with real-time estimates. While the *real-time estimate* for a given year T uses the (vintage) WEO data from year T, the *quasi-real-time estimate* for T is based on the final data (from the fall 2017 WEO) truncated at year T. Differences in both estimates are caused by

---

[10] For the Netherlands and Ireland, a very small number of missing observations in earlier WEO vintages were filled in where earlier as well as later vintages showed same values.

[11] Suppose a different estimation method used for the WEO vintage real time estimate yielded an output gap estimate of -0.5 percent without incorporating expert judgment. If MVF was applied to the same data and yielded an estimate of -1 percent, differing only because of the different filtering method, our approach would attribute +0.5 percent to expert judgment.

data revisions, as data revisions are only embodied in the final data.[12] The first column of Table 4 in sub-section E shows the sample averages of the impact of data revisions on output gap estimates for the 11 baseline countries. The last two rows average across countries, showing that data revisions caused real-time estimates to be generally below final estimates, but on average by less than 0.2 percentage point of potential GDP. The text charts show the examples of Germany and France (see Appendix Figure 3 for other countries).

Real-Time and Quasi Real-Time Estimates

Germany / France

Sources: IMF *WEO* and staff estimates.
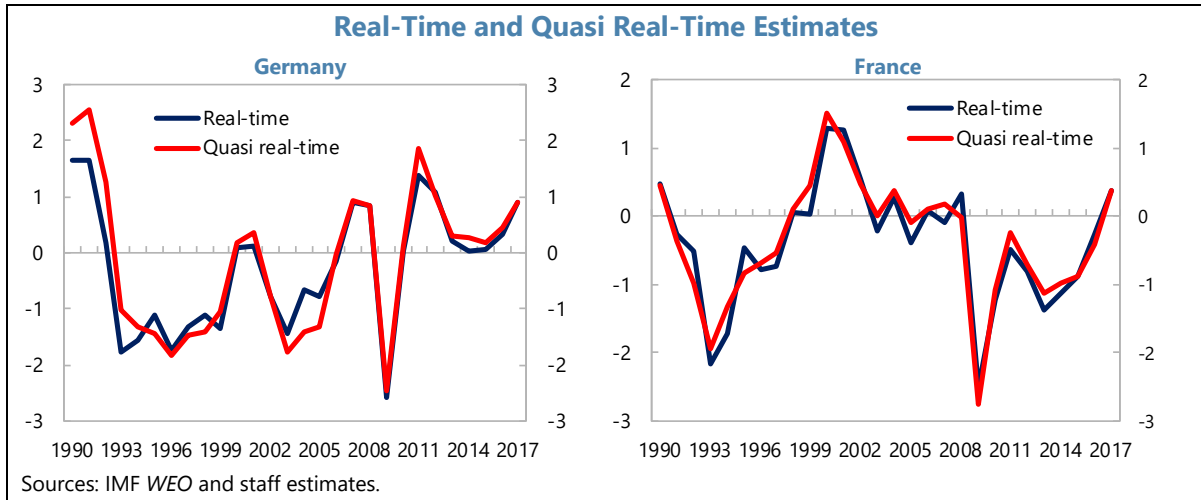
C. Forecast Accuracy

It is well documented that IMF GDP forecasts—as well as those of other institutions—tend to be too optimistic on average (see Box 3 for details). Forecast accuracy matters for all filtering methods because the data sample underlying the estimation does typically not end in year *T* when the estimation takes place. Instead, to improve the end-of-sample stability of the estimates, i.e. to reduce the extent to which a real-time estimate in year *T* is revised when data for *T+1* become available, real-time estimates typically include several years of forecasts. In the case of the MVF real-time estimates presented in this paper, a five-year forecast of the three variables entering the estimation (i.e. output, inflation and unemployment) is used to extend the data sample up to *T+5*.

Impact of Forecast on Today's Potential Estimate
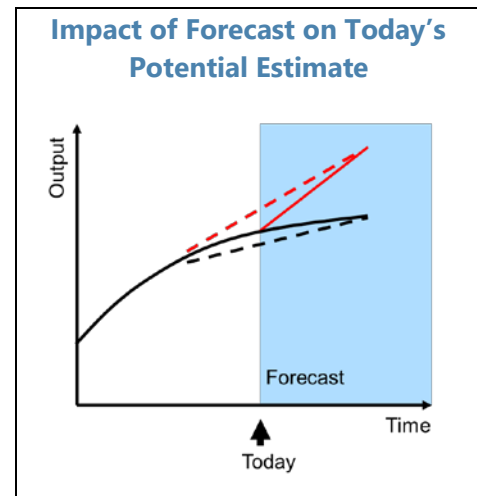
However, extending the data sample comes at the cost of introducing potential spillovers from forecast inaccuracy to inaccuracy in real-time output gap estimates. The reason is that potential output is modelled to be "smooth", so estimated potential in *T* also depends on estimated potential in *T+1* to *T+5,* which is informed by the GDP forecast (and its errors). The text chart illustrates the mechanism. The solid black line indicates the observed output in
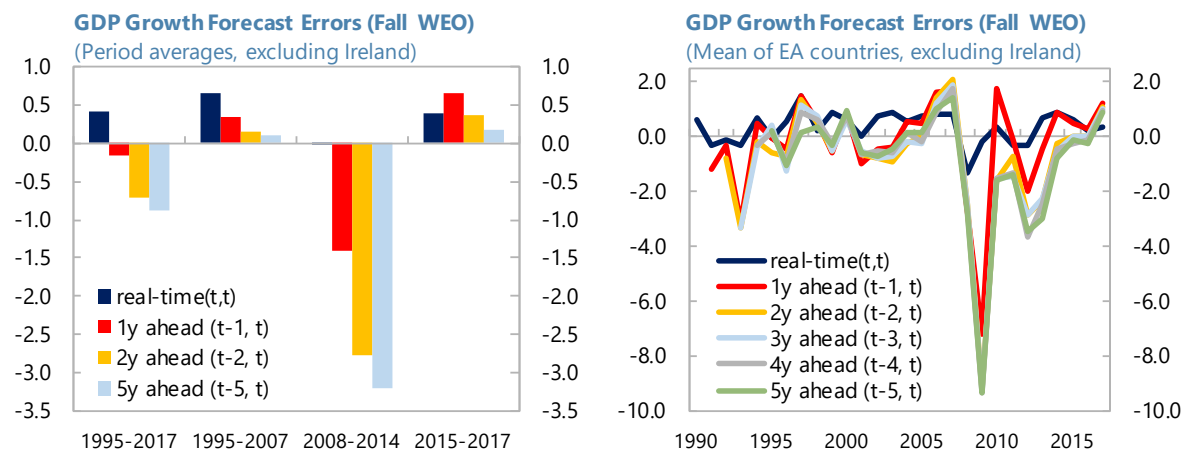
---

[12] To shut off the impact of forecast inaccuracy (discussed in the next subsection), the sample underlying both series is not extended by near-term forecasts.

the historical area of the chart (left of "today") and true future output thereafter, whereas the solid red line symbolizes a too optimistic GDP forecast. Estimated potential is indicated by the dashed lines and, in this simplified example, is thought of as an average between output a few years ago and (forecasted) output a few years ahead. The black dashed line shows estimated potential under a perfect foresight forecast, taking into account the true future output and resulting in a positive output gap estimate for today. When instead forecasted output is taken into account, the estimated potential (red dashed line) exceeds today's output, yielding a negative output gap. In this example, the overly optimistic forecast spills over into a "too negative" output gap estimate (relative to the one obtained under perfect foresight).

Box 3. GDP Growth Forecast Errors Across WEO Vintages

WEO forecasts for the euro area tend to overpredict real GDP growth. The main reason is that baseline forecasts typically do not incorporate the possibility of severe recessions. They are "modal" forecast—and as such display the most likely outcome—not "average" forecasts based on distributional assumptions. Tail events have no quantitative bearings on the baseline forecast. Therefore, when they occur, recessions give rise to large forecast errors.

Building on the IMF IEO report (2014), we analyze GDP growth forecast errors for the 12 euro area countries over the period 1994–2017. We confirm the results established in literature (e.g., Timmermann (2006), De Resende (2014)) that forecasts for nearly all countries and over all forecast horizons (apart from real time, nowcast) exhibit an optimistic bias. However, the bias is statistically significant in less than half of countries in the sample. If the crisis period is excluded from the sample, the size, incidence, and statistical significance of the optimistic bias are reduced substantially.

GDP Growth Forecast Errors (Fall WEO)
(Period averages, excluding Ireland)

- real-time(t,t)
- 1y ahead (t-1, t)
- 2y ahead (t-2, t)
- 5y ahead (t-5, t)

1995-2017   1995-2007   2008-2014   2015-2017

GDP Growth Forecast Errors (Fall WEO)
(Mean of EA countries, excluding Ireland)

- real-time(t,t)
- 1y ahead (t-1, t)
- 2y ahead (t-2, t)
- 3y ahead (t-3, t)
- 4y ahead (t-4, t)
- 5y ahead (t-5, t)

1990   1995   2000   2005   2010   2015

The optimistic bias is stronger in medium-term growth forecasts, where forecast errors are larger and more pronounced, and more countries exhibit statistically significant optimistic bias compared to short-term forecasts. The size of bias also varies with the time-period and countries studied, and mostly stems from a failure to predict recessions. It is, however, noteworthy that the real-time "nowcasts" tend to exhibit a small negative bias, indicating that in real time GDP growth has generally been under-projected.

As GDP growth forecasts are an important input in the estimates of potential output, ex ante optimistic bias in forecasts is translated into overestimating potential output in real time. This optimistic bias could be further exacerbated by an institutional informal rule that require to close the output gap by the end of a five-year forecast horizon (Timmermann, 2006; De Resende, 2014). For example, starting from a negative output gap, in order to close the output gap in the medium-term actual output has to be forecasted to grow rapidly to catch up with potential. As a result, the growth forecast errors can be significant contributors to the output gap revisions.

Before turning to a systematic analysis, we define the following two concepts. First, the *vintage forecast estimate* for year T is based on year-T vintage data and five years ahead WEO vintage forecast. Second, the *perfect-foresight estimate* for year T is different in that the sample is extended by a forecast that is perfect, in the sense that the implied growth rates are correct in hindsight. That is, we compute growth rates for $T+1$ to $T+5$ in the final data (i.e. the fall 2017 WEO) and use those rates to extrapolate the vintage data from year T onwards.[13] The two estimates differ because of the inaccuracy of the vintage forecast, i.e. because of deviations of forecasted growth rates from ex-post observed growth rates.

The second column of Table 4 in sub-section E shows that forecast inaccuracy tends to lower output gap estimates (increase potential output) by around 0.7 percentage points on average, which corroborates the finding of optimistic forecast bias reported in Box 3. Appendix Figure 4 shows time-series charts for 11 countries. The text chart depicts the example of France and shows that the implications of forecast inaccuracy—the difference between the blue and red lines—tends to be larger in the years preceding a sharp slowdown.[14] The closer we get to

France: Vintage and Perfect Foresight Estimates

Sources: IMF *WEO* and staff estimates.

2008, the more prominent is the influence of the crisis period in the 5-year forecast horizon. For the vintage forecast series, the proximity to the crisis has no immediate implications since the crisis was not forecasted and therefore the potential output estimate is too positive and the output gap too small. In the perfect foresight series, in contrast, the crisis is part of the forecast. As we get closer to 2008, the crisis increasingly weighs down on the estimate of potential output and thereby leads to larger positive output gaps.

---
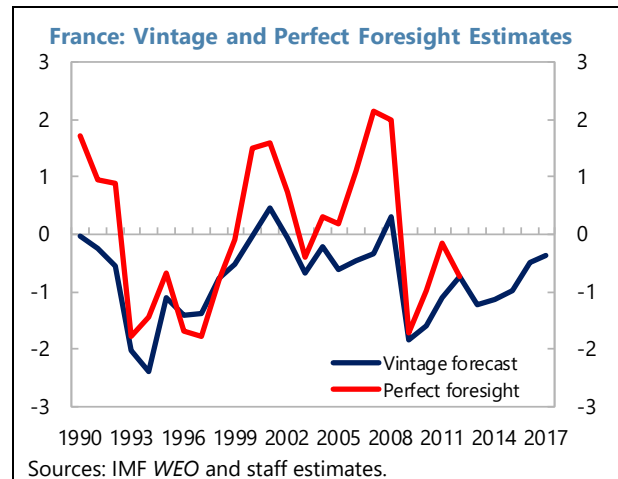
[13] Note that we are using growth rates to construct the perfect-foresight forecast instead of taking the final data for $T+1$ to $T+5$ in levels. The latter would make it impossible to isolate the impact of forecast inaccuracy from the impact of data revisions. Consider a simplified example in which we construct the perfect-foresight estimate for the year 2000. Suppose in final data the output for the year 2000 was revised downward by 1 percent and that output in 2001 grew by 2 percent. If we extended the sample by the level of output (in the final data) from 2001 onwards, we would connect the vintage data for 2000 with the final data for 2001, implying a growth rate of only 1 percent (compared to a true growth rate of 2 percent). Thus, in presence of data revisions, simply taking the final data (in levels) as a forecast means that forecasted growth rate in $T+1$ differs from the growth rate in the final data. The exercise would thus be polluted by the impact of data revisions. Using final-data growth rates to extrapolate the vintage data from T onwards does not suffer from this shortcoming, since the implicit growth rates are ex-post correct when combined with the vintage data.

[14] The perfect foresight series ends in 2012 because its construction requires data for the 5 subsequent years to compute perfect-foresight projections.

E.   The Role of Judgment

Finally, we analyze *judgment*, i.e. deviation of staffs' output gap estimates from the filter results. We define historical judgment as the deviations of WEO real-time estimates from the pure MVF-based vintage estimates. The MVF-based vintage estimates are based on all information available to staff in real time (data as well as forecasts) and can therefore be interpreted as a natural benchmark, an estimate that is free of expert judgment. In other words, the analysis can be interpreted as showing what would staff estimates have been if it had used the MVF filter since the beginning of the sample. Because of the reliance on a particular estimation method (the MVF), however, judgment does not only capture additional information on economic structure (e.g. impact of structural reforms), but also the choice of modelling framework[15] as well as the impact of any other "external" factors, like e.g. over-optimism about convergence and potential growth of EU member countries (see footnote 11).

The third column of Table 4 in the next sub-section shows that the real-time WEO estimates of the output gap are on average about 1 percentage point below the MVF-based vintage estimates. Note that the MVF-based vintage forecast estimates should not be understood as being necessarily superior to WEO real-time estimates because the application of a filter can only complement—but not substitute—staff's assessment, which is based on a much broader information set than the variables entering a filter. However, it is striking that the deviations of WEO real-time from vintage forecast estimates are that large and predominantly negative.

Table 4. MVF Estimates of Real Time Output Gap Biases

	Data revisions		Forecast errors		Judgement		Total bias 1/		WEO revisions 2/	
	Mean	p-value	Mean	p-value	Mean	p-value	Mean	p-value	Mean	p-value
Austria	0.19	0.00	0.44	0.02	0.53	0.00	1.26	0.00	1.39	0.00
Belgium	0.23	0.00	0.27	0.06	0.53	0.00	1.13	0.00	0.87	0.00
France	0.07	0.09	0.71	0.00	1.37	0.00	2.15	0.00	1.45	0.00
Germany	-0.01	0.88	0.64	0.00	0.69	0.00	1.41	0.00	1.31	0.00
Italy	0.03	0.45	1.09	0.00	1.58	0.00	2.68	0.00	1.65	0.00
The Netherlands	0.21	0.02	0.25	0.36	0.45	0.02	0.63	0.00	0.59	0.01
Finland	0.13	0.12	0.63	0.07	1.50	0.00	2.29	0.00	2.49	0.00
Greece	-0.27	0.06	1.41	0.05	-0.93	0.00	0.23	0.73	1.02	0.29
Ireland	0.92	0.00	-0.53	0.47	1.38	0.00	1.64	0.07	-0.24	0.75
Portugal	0.05	0.30	1.13	0.00	0.62	0.00	1.65	0.00	1.60	0.00
Spain	0.37	0.02	0.55	0.05	0.64	0.01	1.75	0.00	2.28	0.00
Unweighted average	0.18	0.00	0.60	0.00	0.76	0.00	1.53	0.00	1.31	0.00
Weighted average	0.11	0.00	0.66	0.00	0.95	0.00	1.74	0.00	1.42	0.00

Sources: IMF *WEO* and staff calculations.
1/ Sum of the impacts of data revisions, forecast errors and judgment for 1994–2012.
2/ Difference between the 2017 Fall WEO and real-time output gap estimates for 1994–2012.
Note: A positive mean indicates the extent of downward bias, causing mean output gap estimates to fall below their respective benchmarks. P-values indicate statistical difference from zero. Weighted average is based on GDP weights.

---

[15] For example, assume that staff used some filtering method to derive a real-time output gap estimate of -2 percent for 1995 without deviating from the filter estimate. Further assume that the MVF, when applied on the historical real-time data, estimates a gap of -1 percent for the same year. The difference in filtering methods will be interpreted as "judgment" while no expert knowledge beyond the use of a different filter was applied. The magnitude of this inaccuracy and its prevalence cannot be determined with precision due to lack of sufficiently detailed information on the specific estimation methods used each year in the past.

F. Comparison to WEO Revisions

The first three columns in Table 4 show the individual contributions of data revision, forecast errors and judgment to the total bias summarized in column 4. Column 5 shows the actual revisions to WEO output gap estimates (average final estimate minus average real-time estimate). Comparing the last two columns shows that the total impact of the three sources of uncertainty is close to the size of WEO revisions, with an unweighted average difference of 0.2 percentage points of potential GDP. According to this exercise, just below half of WEO revisions can be attributed to forecast errors and the remaining half to judgment, with data revisions having a minor role.[16]

V. RELEVANCE TO MONETARY POLICY AND INFLATION FORECAST

Monetary policy reaction functions usually include a measure of slack—that is most often captured by the output gap—and a measure of inflation expectations, with various leads and lags to denote both the forward-looking nature of policy and a desire for gradual adjustment (through interest rate persistence). Therefore, real-time biases in output gap estimates could lead to policy errors either through its direct impact on the policy instrument (Orphanides, 2003) or through lower predictability of inflation (Orphanides and van Norden, 2005).[17]

In this section we test whether the WEO output gap is useful in explaining inflation. Tables 5 and 6 show basic Phillips curve relationships in real time and in the final 2017 WEO vintage, respectively, for different sample sizes and time periods. While the Phillips curve relationships seems to be strongly confirmed based on final 2017 WEO data, the coefficient for the real-time output gap is small and generally non-significant statistically.[18] The real time WEO output gap does not seem to provide additional information over and above what is already contained in past inflation and inflation expectations. It is therefore of limited use as a basis for advice on monetary policy.

Only when the sample size is reduced to the years when countries entered the monetary union can the real-time output gap be statistically associated with inflation, in some specifications. Nevertheless, such association is rare and not robust, and even for this time period, the revisions to real-time output gaps are substantial. Notwithstanding large revisions,

---

[16] Note that since the three sources of biases are derived from strict application of the rolling-window MVF on the real time data, whereas revisions are calculated vis-à-vis 2017 WEO final data estimates that can still incorporate expert judgment, the sum of the three biases can differ from WEO revisions. France, Italy and Ireland stand out with higher MVF biases compared to WEO revisions.

[17] Findings on the usefulness of real time output gap estimates for forecasting inflation are mixed. Orphanides (2000, 2005) suggested that mis-measuring the output gap in real time adversely contributed to economic policy decisions in the past. Champagne and others (2018) also find some deterioration in forecast performance when inflation projections are conditioned on real time rather than on final estimates of the gap. Marcellino and Musso (2011) point to marginal usefulness of real-time output gap estimates for forecasting inflation in the euro area both in the short term (one quarter and one year ahead) and the medium term (two and three years ahead).

[18] This result is in line with similar findings for the US in Orphanides (2005). Recent work by McLeay and Tenreyro (2019) suggests that accounting for negative correlation between inflation and output gap implied by an optimal forward-looking targeting rule helps to identify a positively sloping Phillips curve between inflation and the output gap.

simple correlation coefficients between the changes in real-time and final output gap estimates are positive at 67 percent for 11 baseline countries and 28 percent for all countries, and highly significant. Thus, real-time output gaps can to some extent be informative on the direction or turning points of final output gap and hence inflation. This echoes earlier findings by Tosetto (2008) who points out that notwithstanding large and persistent revisions to the initial output gaps published by the OECD, preliminary estimates are strongly correlated with the successive ones, can serve as useful predictors of the latter, and in this respect may still contain useful information for future inflation and monetary policy.

The lack of a "real-time" Phillips curve and negligible ex-post revisions to WEO inflation forecasts mean that the real-time output gap bias is not reflected in inflation forecasts, mirroring the important disconnect between inflation (fluctuating around 2 percent, close to the euro area inflation expectations) and persistently negative real-time output gaps (see Section III.B). Due to high inflation persistence in Europe and small revisions, inflation in real time is a useful predictor of final inflation and informative for monetary policy.[19] To the extent that real-time output gap bias leads to a loss in inflation predictability, or policy rules and decisions rely on real time estimates of cyclical position, monetary policy can still suffer from information uncertainty.[20]

Table 5. Basic Real-Time Phillips Curve Regressions

	Real-time inflation									
	Full sample		1994-2012		2000-12		2000-12 euro area		2000-07 euro area	2008-12 euro area
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Real-time output gap	0.062	-0.012	0.012	-0.014	0.048	-0.021	0.186***	0.073	0.025	0.065
	(0.04)	(0.02)	(0.03)	(0.03)	(0.06)	(0.04)	(0.06)	(0.05)	(0.04)	(0.11)
Real-time vintage inflation (t-1)		0.406***		0.398***		0.319***		0.346***	0.187	0.140
		(0.05)		(0.06)		(0.08)		(0.08)	(0.12)	(0.11)
Real-time vintage inflation expectations (t+2)		0.755***		0.739***		0.687***		0.498*	1.097***	0.244
		(0.16)		(0.20)		(0.25)		(0.26)	(0.25)	(0.35)
Constant	2.250***	-0.264	2.926***	-0.242	2.619***	0.226	2.058***	-1.449***	0.024	-0.353
	(0.27)	(0.36)	(0.28)	(0.42)	(0.27)	(0.57)	(0.25)	(0.35)	(0.33)	(0.72)
R-squared	0.64	0.85	0.60	0.83	0.64	0.77	0.70	0.79	0.83	0.82
Adjusted R-squared	0.59	0.82	0.53	0.80	0.57	0.72	0.64	0.74	0.78	0.75
Observations	351	313	256	256	186	186	175	175	94	81

Source: IMF WEO and staff estiamtes.

Note: Standard errors in parentheses. ***, **, and * denote significance at 1, 5, and 10 percent, respectively. All regressions include time and fixed effects.

---

[19] This feature would only remain under stable inflation expectations.

[20] This statement is conditioned on assuming that the output gap at the euro area level is related to the individual estimates in the panel.

Table 6. Basic Ex-Post Phillips Curve Regressions

| | 2017 WEO Inflation | | | | | | | | | |
| | Full sample | | 1994-2012 | | 2000-12 | | 2000-12 euro area | | 2000-07 euro area | 2008-12 euro area |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2017 WEO output gap	0.218***	0.142***	0.246***	0.158***	0.218***	0.216***	0.130***	0.077*	-0.060	0.162**
	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.06)	(0.06)
2017 WEO inflation (t-1)		0.543***		0.534***		0.535***		0.389***	0.349***	-0.091
		(0.04)		(0.04)		(0.09)		(0.08)	(0.11)	(0.11)
2017 WEO vintage inflation expectations (t+2)		0.156**		0.132*		-0.040		0.159	0.080	-0.291*
		(0.07)		(0.08)		(0.10)		(0.10)	(0.09)	(0.17)
Constant	4.095***	0.919***	0.147	0.969***	0.993**	0.007	1.404***	0.852***	1.027***	3.383***
	(0.80)	(0.28)	(0.58)	(0.30)	(0.42)	(0.42)	(0.25)	(0.26)	(0.37)	(0.46)
R-squared	0.47	0.85	0.46	0.84	0.53	0.69	0.64	0.70	0.68	0.83
Adjusted R-squared	0.42	0.83	0.40	0.82	0.46	0.64	0.56	0.64	0.59	0.76
Observations	438	398	343	341	245	245	177	177	96	81

Source: IMF WEO and staff estiamtes.

Note: Standard errors in parentheses. ***, **, and * denote significance at 1, 5, and 10 percent, respectively. All regressions include time and fixed effects.

VI. RELEVANCE TO FISCAL POLICY OUTCOMES

Persistent bias in real-time output gap estimates could have implications for fiscal policy. Output gaps are typically used to separate fiscal balances into their cyclical and structural components. The latter is supposed to provide a measure of the government's underlying fiscal policy stance—cleaned from temporary cyclical influence—and help steer the desirable setting of fiscal policy given the authorities' medium-to-long-term debt consolidation objectives (which in the European stability and growth pact framework is formalized in terms of the medium-term objective). If real-time output gaps fail to properly disentangle trend from cycle, then the long-term public debt implications of a given deficit would be inaccurately estimated. For example, a negative output gap bias in real time would result in an excessively optimistic view of the long-term level of output and fiscal revenues.
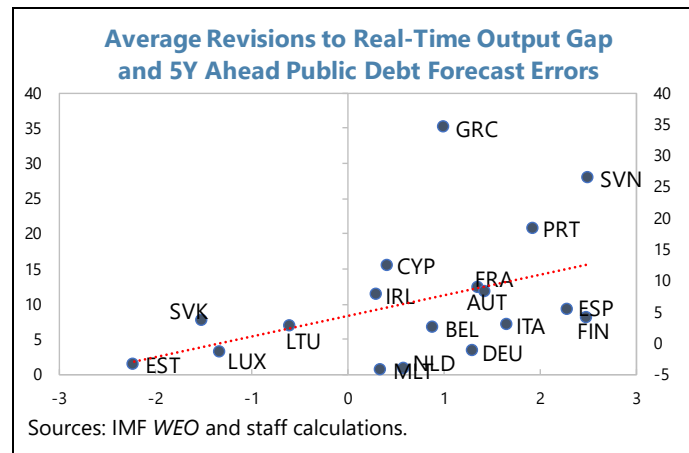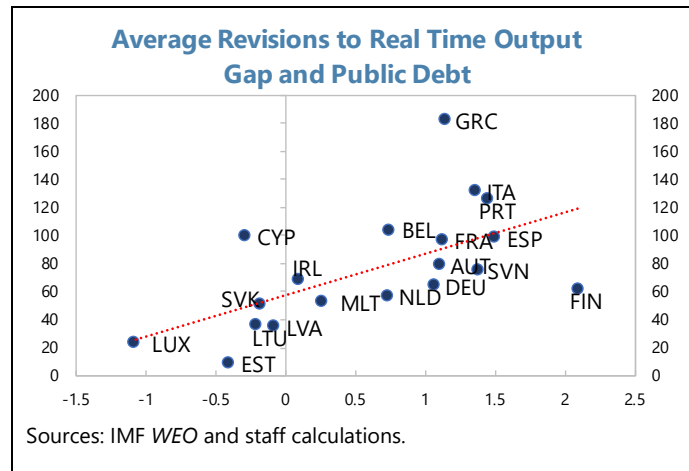
To the extent that the estimates of the output gap and structural fiscal balance play a role in calibrating fiscal policy, over-optimism about long-term growth could translate into excessive deficit and debt buildup. Simulations by Ley and Misch (2014) based on WEO data on output gap revisions for 175 countries and over 17 years show that in more than one-fifth of cases the implied revisions of the overall and structural fiscal balances exceed 1 percent of GDP, possibly leading to unplanned debt accumulation or surprise debt reduction. They caution against taking real-time estimates of structural balances at face value and recommend building safety margins in the fiscal policy targets.

To fix ideas about the magnitudes involved, consider the following highly simplified thought experiment. Assume that countries' fiscal authorities had relied on real-time estimates of the output gap to calibrate fiscal policy, with similar biases as the WEO estimates in terms of the magnitudes and signs, two conditions that would require further verification. We can then translate the real time output gap bias into a cyclical component of fiscal balances, using an approach similar to Kempkes (2014). The resulting bias in fiscal policy would range from 0.7 to

0.9 percent of potential GDP annually.[21] Year after year overestimation of potential would thus imply significant unplanned deficits and large debt buildup over a 19-year baseline sample.[22]

In a similar exercise, Eyraud and others (2018) and Eyraud and Wu (2015) compare ex post estimates of the output gap with real time estimates in the stability programs prepared annually by the euro area country authorities for the European Commission. For the period 2003–16 Eyraud and others (2018) find that the output gap was underestimated in real time by 1.3 percentage points of GDP on average. Assuming an elasticity of revenue to output of 1, and elasticity of expenditure to output of 0, and an average expenditure to GDP ratio of 45 percent the underestimation of output gap in real time would imply an overestimation of cyclically adjusted balances by 0.5 percentage point on average per year. Given that euro area countries are subject to cyclically-adjusted balance rule, negative bias in the authorities' real time output gap estimates may have led to unplanned deficits.

Average Revisions to Real Time Output Gap and Public Debt

Sources: IMF *WEO* and staff calculations.

Drawing on the WEO database over the period 1994 to 2017, we find a positive association between the level of the debt-to-GDP ratio in 2017 and the average output gap revisions over the past 24 years (text chart). High-debt euro area countries have seen larger upward revisions to WEO real-time output gaps, as illustrated in the text chart. As shown in Appendix Table 3, the average revision in higher-debt countries amounts to 1.1 percent of potential GDP, while real-time estimates in lower-debt countries are on average only revised by 0.5 percent, thus closing the difference between the average output gap estimates for both country groups

Average Revisions to Real-Time Output Gap and 5Y Ahead Public Debt Forecast Errors

Sources: IMF *WEO* and staff calculations.

---

[21] Depending on whether the bias is computed based on WEO output gap revisions or differences between MVF real time simulations and real time output gap estimates (see Appendix Table 2).

[22] In our exercise the average increase in debt over 19 years is around 15 percent of GDP and can reach close to 25 percent of GDP in countries with larger bias. The calculation assumes that the real-time output gap bias and the corresponding cyclical fiscal component (obtained by applying the budgetary semi-elasticities estimated by Mourre and others, 2014) mechanically translates into a structural balance bias, which is multiplied by the number of years over which the average bias is computed to obtain the debt implications. This is a very rough calculation that neither takes into account interest rates nor the sequence of deficit nor feedbacks from fiscal policy to output.

in the final vintage. Similarly, WEO output gap revisions over the period 1994–2012 are positively associated with WEO 5-year ahead forecast errors for the public debt ratio. [23]

Box 4 builds on the cross-sectional correlations between average output gap revisions over time on the one hand and the public debt ratio in 2017 as well as forecast errors for public debt on the other. Specifically, it investigates the correlation between output gap revisions and primary fiscal balances, while conditioning on a variety of variables that can affect these fiscal balances. In so doing, it takes advantage of both the time series and cross-sectional dimensions of the data. As expected, the association between the final output gap estimate and the primary fiscal balances is positive: the better the economy is doing, the higher is the fiscal balance. The association between revisions to real time output gaps estimates and primary fiscal balances is, however, strongly negative: larger upward revisions to WEO output gaps are associated with lower primary fiscal balances. In other words, the countries for which WEO forecasters have seen too much economic slack and thus too much potential for the economy to grow in the future also tend to be the countries that have seen lower primary fiscal balances.

The excess slack in WEO real-time estimates are related to both overly optimistic forecasts as well as overoptimistic potential growth incorporated through judgment. Both are associated with lower primary balances (Box 4). While the former is a corollary of difficulties with projecting recessions (see Section IV), the origin of the latter is not clearly identifiable. It could reflect overestimates of the impact of structural and fiscal reforms on potential output, but also many other factors. The results suggest that such over-optimism is strongly associated with lower than expected primary balances and debt accumulation.

To conclude, we derive the following insights from our analysis of output gaps and fiscal outcomes: (i) WEO forecasters have seen more slack in real time than they considered warranted ex post—their real time output gap estimates were persistently negative; (ii) this has been more pronounced for countries with high public debt; (iii) where it has been more pronounced, it has also been associated with (a) larger WEO forecast mistakes with respect to public debt ratios, and (b) lower primary balances. Simple back-of-the-envelope calculation shows that the unexpected debt accumulation due to real time output gap biases could potentially be substantial. Similarly, evidence presented by Eyraud and others (2018) and Eyraud and Wu (2015) suggests that similar large negative bias is present in country authorities' real time output gap estimates that may have led to unplanned deficits under a structural fiscal balance rule. Further investigation is required to establish whether and to what extent persistently negative real time output gap estimates have lured country authorities into adopting looser fiscal policies than they would have wanted.

The uncertainty associated with estimating an unobservable concept such as potential GDP calls for fiscal policy prudence. This is especially relevant for high debt countries which lack fiscal space, as an overestimation of potential output could increase the risk of unsustainable fiscal outcomes. In practice, the acknowledgment of high uncertainty would suggest stepping away from point estimates and relying more on the whole distribution of output gap estimates. The dispersion and skewness of the estimated distribution can provide useful

---

[23] General government debt and fiscal balances have been affected by financial sector support measures, mostly at times of GFC and European sovereign debt crises when forecast errors were large.

information to policy-makers when calibrating fiscal policies. More generally, our analysis cautions against extensive focus on real-time output gaps when calibrating fiscal policy. Eyraud and others (2018) point out that expenditure rules that allow automatic stabilizers to work, although more sensitive to initial conditions, can be more resilient to measurement errors. Andrle and others (2015) show with counterfactual simulations for France and Italy that the differences between real time and ex post outcomes of public debt from 2001 onwards would have been significantly smaller under the expenditure rule than under cyclically adjusted balance rule.

Box 4. Fiscal Outcomes and Real-Time Output Gap Bias

To study the relationship between fiscal outcomes and WEO output gap revisions we adapt a standard framework of fiscal reaction functions popularized by Bohn (1998, 2008). Table 7 reports the results of regression analysis that relates primary fiscal balances with the lagged primary balance (persistence), lagged public debt and detrended real government consumption (optimizing conditions due to Barro, 1979) as well as the WEO output gap and revisions to real time WEO output gap (the primary channel of interest). The latter is also broken down by MVF simulated sources of real-time bias (derived in the previous section). The regressions further control for a standard set of institutional and economic variables along with 5-year ahead debt forecast errors to capture for any possible feedback of future changes in fiscal policies on output gap revisions as well as real-time inflation and unemployment gaps to net out residual cyclical effects.[1]

Our choice of estimator is the bias-corrected Least Squares Dummy Variable (LSDV) dynamic panel estimator by Bruno (2005) to correct for the "dynamic panel bias" in samples with small N that have been applied in similar framework of primary balance regressions (see, for example, Debrun and Kinda, 2013). We do not consider issues of endogeneity and simultanety in fiscal policy in much depth because we are mostly interested in checking whether the insights from the bivariate, cross-sectional correlations between output gaps revisions and fiscal outcomes discussed above are robust to further testing in a multivariate, time series framework. We find that this is the case.

The coefficient on real time output gap revisions is statistically highly significant and economically almost as large as the one on the final output gap. Estimated negative coefficients in columns 2–5 indicate that predominantly positive (upwards) revisions to WEO real-time output gaps averaging to 1.3–1.4 percent of potential GDP (Table 7) are associated with lower primary balance estimates in the order of 0.8–1.2 percent of GDP annually. The regression results also indicate that the association is broad-based, working through both the MVF simulated effect of forecast errors and judgment with similar coefficients. While the data revisions tend to have a statistically significant positive coefficient, likely indicating that revisions to real-time GDP achieve better consistency with fiscal policy outcomes, the economic magnitude of such revisions is small and driven mostly by one country in the sample.

The estimated associations could in principle be driven by a common factor: countries where GDP growth has disappointed can end up with both upward revisions to output gaps and lower deficits, leading to higher debt-to-GDP ratios. In this regard, the positive coefficient on the current account balance, besides accounting for twin deficits, is expected to capture correlated international spillovers (see Checherita-Westphal and Žďárek, 2017) while inflation and unemployment gaps control for the impact of the business cycles on the deficit. The baseline bias-corrected LSDV estimates are also robust to time effects that capture a basic form of cross-country dependence (see Appendix Table 4). Estimated coefficients of other determinants of primary fiscal balances are broadly in line with earlier studies. The highly significant autoregressive component indicates persistence in primary balances. The positive coefficient on lagged public debt may capture that governments with higher public debt try to achieve higher primary balances (the "weak sustainability condition"). Results are robust to sample size and real-time debt (results not reported).

---

[1] Shifts in the cyclical and trend components are captured by "inflation gap" (inflation deviation from its long-term average), the "unemployment gap" (deviation of unemployment rate from NAIRU), and a "NAIRU gap" (deviation of estimated NAIRU from long-term average unemployment rate). The two former capture mostly cyclical movements while the NAIRU gap captures shifts in (common) trends (see Section III.B for discussion).

Table 7. Primary Balance Regressions

	Primary fiscal balance (percent of GDP)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Primary balance (t-1)	0.641*** (0.06)	0.498*** (0.06)	0.393*** (0.06)	0.473*** (0.06)	0.406*** (0.06)	0.514*** (0.06)	0.527*** (0.06)	0.518*** (0.06)
Output gap (final vintage)	0.327*** (0.09)	0.801*** (0.12)	0.767*** (0.15)			0.154 (0.13)	0.335** (0.15)	0.196 (0.13)
Revision to real time output gap against 2017 vintage		-0.714*** (0.14)	-0.849*** (0.16)					
Output gap (t+5 vintage)				0.834*** (0.11)	0.832*** (0.14)			
Revision to real time output gap agains (t+5) vintage				-0.612*** (0.15)	-0.745*** (0.17)			
Data revisions (MVF)						1.023*** (0.37)		
Forecast error (MVF)							-0.444* (0.24)	
Judgement (MVF)								-0.490** (0.24)
Public debt (t-1)	0.039** (0.02)	0.035** (0.02)	0.082*** (0.02)	0.041** (0.02)	0.079*** (0.02)	0.082*** (0.02)	0.086*** (0.02)	0.085*** (0.02)
Current account (in percent of GDP)	0.254*** (0.09)	0.217** (0.09)	0.277*** (0.09)	0.241*** (0.08)	0.281*** (0.09)	0.255*** (0.09)	0.234** (0.09)	0.276*** (0.09)
Government real consumption gap	-0.098 (0.06)	-0.100* (0.06)	-0.056 (0.06)	-0.099* (0.06)	-0.061 (0.06)	-0.026 (0.06)	-0.012 (0.06)	-0.045 (0.06)
Trade openness	-0.060*** (0.02)	-0.084*** (0.02)	-0.111*** (0.02)	-0.089*** (0.02)	-0.107*** (0.02)	-0.073*** (0.02)	-0.074*** (0.02)	-0.096*** (0.02)
IMF arrangement	-0.334 (1.00)	0.583 (0.93)	0.318 (0.87)	0.496 (0.91)	0.232 (0.87)	-0.156 (0.94)	-0.113 (0.96)	-0.313 (0.94)
Election year	-0.261 (0.44)	-0.177 (0.41)	-0.222 (0.38)	-0.186 (0.40)	-0.239 (0.37)	-0.139 (0.40)	-0.172 (0.40)	-0.210 (0.40)
5Y ahead public debt forecast error			0.023 (0.01)		0.014 (0.01)	0.022 (0.02)	0.022 (0.02)	0.010 (0.02)
Inflation gap			0.656*** (0.16)		0.656*** (0.15)	0.784*** (0.17)	0.783*** (0.17)	0.783*** (0.17)
Unemployment gap			0.004 (0.17)		0.065 (0.17)	-0.228 (0.16)	-0.227 (0.17)	-0.290* (0.16)
NAIRU gap			-0.294** (0.14)		-0.251* (0.14)	-0.160 (0.15)	-0.190 (0.15)	-0.284* (0.17)
Observations	195	195	195	195	195	195	195	195

Sources: IMF *WEO* and staff estimates.

Note: Standard errors in parentheses. ***, **, and * denote significance at 1, 5, and 10 percent, respectively. Baseline sample for 11 larger euro area countries for the period 1994–2012. All regressions are estimated with the bias-corrected LSDV dynamic panel data estimator with bootstrapped standard errors.

VII. CONCLUSIONS

The output gap, while unobservable, is typically considered an important input for policymaking. Policymakers need to be able to assess the underlying growth rate of the economy to separate trend from cycle and optimally calibrate the stance of monetary and fiscal policies. Failure to do so may result in excessive inflation or debt buildup as monetary and fiscal policy become pro-cyclical. However, measuring the output gap in real time is very difficult since information is scarce and hard to interpret, structural shifts may not be immediately visible, and recessions are difficult to forecast. A large empirical literature has shown that output gaps tend to be revised over time, and the revisions may be large, to the point that what looks like an optimal countercyclical policy in real time may appear pro-cyclical in hindsight.

In this paper we audit the WEO real-time output gap estimates for countries in the euro area against a set of desirable properties. We show that revisions of WEO real-time output gap estimates for the main euro area countries are large—especially before the GFC—and systematically upwards. Consequently, staff's real-time output gap estimates are persistently downward biased, sometimes extending over more than a decade. Counterfactual simulations show that over-estimation of the economy's productive capacity and over-optimistic growth forecasts are the main drivers of the negative bias. The forecast errors primarily occur because WEO forecasters aim at predicting the mode of the growth distribution which, in the presence of downwardly skewed risks, exceeds average growth. The judgment component of the negative bias argues for caution in incorporating off-model information (such as structural or fiscal reform impact) and favors filtering methods that are not too sensitive to the end of the sample.

Such uncertainties over real-time output gap estimates can have important consequences on policy to the extent it is informed by these estimates. We find that in countries where slack (and potential growth) is overestimated in the WEO data, primary fiscal balances tend to be lower, public debt ratios higher and increase faster than projected. It is thus possible that the biases in output gap estimates reflected an underappreciation of a slowdown in potential growth, which WEO forecasters and policymakers have come to understand only gradually. What were deemed cyclical revenue losses became structural losses, leading to a large accumulation of debt. Available evidence suggests that the national authorities' real-time output gap estimates have suffered from a similar negative bias. To the extent that those estimates played a role in calibrating fiscal policy, over-optimism about long-term growth could have contributed to excessive deficits and larger debt buildup.

We also find, in line with earlier literature, that real-time output gaps are not useful to predict inflation, as evidenced by the apparent lack of "real time" Phillips curve. Real-time output gaps at best contain information on directional changes in business cycles and inflation. Because inflation is persistent and not revised, inflation in real time remains a useful indicator to forecast inflation.

To minimize the occurrence of large measurement errors in the future, staff has moved towards a multi-pronged approach to the estimation of output gaps (IMF, 2015). The potential output is estimated using a multivariate filter (MVF) that—thanks to the economic

structure imposed during estimation—yields real-time potential output estimates that are more in line with desirable economic properties of output gaps and less prone to revisions. As the model structure imposed on the data may not capture important structural changes in the individual economies, the approach still allows to introduce expert judgment. The latter is added during the filtering procedure and is based on growth accounting via a production function approach.

Several options are currently being studied to further improve the reliability of real-time output gap estimates and their usefulness for policymaking. First, increasing the number of indicators (capacity utilization, investment to GDP, labor vacancy rates, etc.) may improve the precision of the MVF and reduce the size of ex-post revisions (Turner and others, 2016). Second, incorporating some measure of the forecast distribution asymmetry in the central forecast may reduce forecast errors and ex-post revisions. One possibility would be to rely on the mean instead of the mode of forecasts. As shown in this paper, failure to predict recessions has been an important driver of a systematic downward bias in real time output gap estimations. Third, complementary measures of the output gap can help hedge against first order mistakes in real time. IMF (2015) discusses a related but different concept of "sustainable" output: the GDP level that an economy can sustainably produce over the medium term in the absence of imbalances. This approach yields a complementary measure of potential output that can provide timely and useful information for real-time policymaking, especially when the two measures diverge significantly. Finally, policymakers could consider building in safety margins or other robust measures to shield against real time biases. This would require building buffers during upswings, increasing the available resources for effective counter-cyclical fiscal policy in downturns. All in all, our paper cautions against an exclusive focus on output gaps estimates in real time when calibrating counter-cyclical policy and stresses the importance of an encompassing approach.

REFERENCES

Abdih, Y., Lin, L. and A.-C. Paret, 2018, "Understanding Euro Area Inflation Dynamics: Why So Low for So Long?" IMF Working Paper WP/18/188.

Ademmer, M., Boysen-Hogrefe, J., Carstensen, K., Hauber, P., Jannsen, N., Kooths, S., Rossian, T. and U. Stolzenburg, 2019, "Schätzung von Produktionspotenzial und-lücke: Eine Analyse des EU-Verfahrens und mögliche Verbesserungen*,"* Kieler Beiträge zur Wirtschaftspolitik, No. 19.

Alichi, A., Avetisyan, H., Laxton, D., Mkhatrishvili, S., Nurbekyan, A., Torosyan, L. and H. Wang, 2019, "Multivariate Filter Estimation of Potential Output for the United States: An Extension with Labor Market Hysteresis," IMF Working Paper (*forthcoming*).

Álvarez, L.J. and A. Gómez-Loscos, 2018, "A menu on output gap estimation methods," *Journal of Policy Modeling*, (40)4, pp. 827-850.

Aiyar, S. and Voigts, S., 2019, "The negative mean output gap", *IMF Working Paper*, forthcoming.

Andrle, M., Bluedorn, J., Eyraud, L., Kinda, T., Koeva Brooks, P., Schwartz, G. and A. Weber, 2015, "Reforming Fiscal Governance in the European Union," IMF Staff Discussion Note, SDN/15/09.

Ball, L., 1999, "Aggregate Demand and Long-Run Unemployment", *Brookings Papers on Economic Activity*, vol. 2, pp. 189-251.

Barro, R., 1979, "On the Determination of Public Debt," *Journal of Political Economy*, 87(5), pp. 940-971.

Bashar, O., 2011, "On the permanent effect of an aggregate demand shock: Evidence from the G-7 countries", *Economic Modelling*, 28(3), pp. 1374-1382.

Benes. J., Clinton, K., Garcia-Saltos, R., Johnson, M., Laxton, D., Manchev, P. and T. Matheson, 2010, "Estimating Potential Output with a Multivariate Filer," IMF Working Paper WP/10/285.

Benigno, P. and L.A. Ricci, 2011, "The Inflation-Output Trade-Off with Downward Wage Rigidities," *American Economic Review*, 101(4), pp. 1436-1466.

Berger, H., Dowling, T., Lanau, S., Lian, W., Mrkaic, M., Rabanal, P. and M.T. Sanjani, 2015, "Steady as She Goes–Estimating Potential Output During Financial "Booms and Busts"," IMF Working Paper WP/15/233.

Blagrave, P., Garcia-Santos, R., Laxton, D. and F. Zhang, 2015, "A Simple Multivariate Filter for Estimating Potential Output," IMF Working Paper WP/15/79.

Blanchard, O., Cerutti, E. and L. Summers, 2015, "Inflation and Activity – Two Explorations and their Monetary Policy Implications," IMF Working Paper WP/15/230.

Blanchard, O., and L. Summers, 1986, "Hysteresis and the European Unemployment Problem," *NBER Macroeconomics Annual*, vol. 1, pp. 15-90.

Bohn, H., 2008, "The Sustainability of Fiscal Policy in the United States." In: Neck, R., and Sturm, J.E., Eds., Sustainability of Public Debt, MIT Press, Massachusetts, 15-49.

Bohn, H., 1998, "The Behavior of U.S. Public Debt and Deficits," *Quarterly Journal of Economics*, 113(3), pp. 949-63.

Borio, C., Disyatat, P. and M. Juselius, 2014, "A parsimonious approach to incorporating economic information in measures of potential output," BIS Working Paper No 442.

Bruno, G.S.F., 2005, "Approximating the bias of the LSDV estimator for dynamic unbalanced panel data models," *Economic Letters*, 87(3), pp. 361-366.

Camba-Mendez, G. and D. Rodriguez-Palenzuela, 2003, "Assessment criteria for output gap estimates," *Economic modelling*, 20(3), pp. 529-562.

Champagne, J., Poulin-Bellisle, G. and Sekkel, R., 2018, "The Real-Time Properties of the Bank of Canada's Staff Output Gap Estimates," *Journal of Money, Credit and Banking*, *50*(6), pp.1167-1188.

Chatterjee, P. and S.B. Nowak, 2016, "Forecast Errors and Uncertainty Shocks," IMF Working Paper WP/16/228.

Checherita-Westphal, C. and V. Žďárek, 2017, "Fiscal Reaction Function and fiscal fatigue: evidence for the euro area," ECB Working Paper 2036.

Coibion, O., Gorodnichenko, Y, and M. Ulate, 2017, "The Cyclical Sensitivity in Estimates of Potential Output", NBER Working Paper No. 23580.

De Resende, C., 2014, "An Assessment of IMF Medium-Term Forecasts of GDP Growth," IEO Background Paper No. BP/14/01 (Washington: Independent Evaluation Office of the IMF).

Debrun, X. and T. Kinda, 2013, "That Squeezing Feeling: The Interest Burden and Public Debt Stabilization," IMF Working Paper WP/13/125.

ECB, 2011, "Recent evidence on the uncertainty surrounding real-time estimates of the euro area output gap," Monthly Bulletin, Economic and Monetary Developments, November 2011, Box 5.

ECB, 2005, "The (un)reliability of output-gap estimates in real time," Monthly Bulletin, Economic and Monetary Developments, February 2005, Chapter 4, Box 5, pp. 43-45.

Edge, R.M. and J.B. Rudd, 2016, "Real-Time Properties of the Federal Reserve's Output Gap," *The Review of Economics and Statistics*, 98(4), pp. 785-791.

Eyraud, L., Debrun, X., Hodge, A., Lledó, V. and C. Pattillo, 2018, "Second-Generation Fiscal Rules: Balancing Simplicity, Flexibility, and Enforceability," IMF Staff Discussion Note, SDN/18/04.

Eyraud, L. and T. Wu, 2015, "Playing by the Rules: Reforming Fiscal Governance in Europe", IMF Working Paper, WP/15/67.

Grigoli, F., Herman, A., Swiston, A. and G. Di Bella, 2015, "Output Gap Uncertainty and Real-Time Monetary Policy," IMF Working paper WP/15/14.

Hamilton, J.D. (2018), "Why You Should Never Use the Hodrick-Prescott Filter," *The Review of Economics and Statistics*, 100(5), pp. 831-843.

Hernández de Cos, P., Lacuesta, A. and E. Moral-Benito, 2016, "An exploration of real-time revisions of output gap estimates across European countries," Banco de España Occasional Paper No. 1605.

Hodrick, R.J. and E.C. Prescott, 1997, "Postwar U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit and Banking*, 29(1), pp. 1-16.

Hristov, A., Raciborski, R. and V. Vandermeulen, 2017, "Assessment of the plausibility of the output gap estimates," *Economic Brief*, *23*.

IEO, I., 2014, "IMF Forecasts: Process, Quality, and Country Perspectives," Independent Evaluation Office of the International Monetary Fund.

IMF, 2015, "Where are we headed? Perspectives on potential output," *World Economic Outlook, April 2015: Uneven growth, short-and long-term factors*, pp. 69-110.

Kempkes, G., 2014, "Cyclical adjustment in fiscal rules: Some evidence on real-time bias for EU-15 countries," *FinanzArchiv: Public Finance Analysis*, 70(2), pp. 278-315.

Kumhof, M., Laxton, D., Muir, D. and S. Mursula, 2010, "The Global Integrated Monetary and Fiscal Model (GIMF) – Theoretical Structure," IMF Working Paper WP/10/34.

Ley, E. and F. Misch, 2014, "Real-Time Macro Monitoring and Fiscal Policy," ZEW Discussion Paper No. 14-122.

Marcellino, M. and Musso, A., 2011, "The reliability of real-time estimates of the euro area output gap," *Economic Modelling*, 28(4), pp. 1842-1856.

McLeay, M. and S. Tenreyro, 2019, "Optimal Inflation and the Identification of the Phillips Curve," forthcoming in Eichenbaum, M.S., Hurst, E. and J.A. Parker ed. "*NBER Macroeconomics Annual*," vol. 34.

Mourre, G., Astarita, C. and S. Princen, 2014, "Adjusting the budget balance for the business cycle: the EU methodology," in European Economy, Economic Papers 536.

Nelson, E. and Nikolov, K., 2003, "UK inflation in the 1970s and 1980s: the role of output gap mismeasurement," *Journal of Economics and Business*, 55(4), pp. 353-370.

Nordhaus, W.D., 1987, "Forecasting efficiency: concepts and applications," *The Review of Economics and Statistics*, 69(4), pp. 667-674.

Okun, A.M., 1962, "Potential GNP: Its Measurement and Significance," in Proceedings of the Business and Economic Statistics Section, pp. 98-104 (Washington: American Statistical Association).

Orphanides, A., 2003, "The Quest for Prosperity without Inflation," *Journal of Monetary Economics*, 50(3), pp. 633-663.

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

Orphanides, A. and S. Van Norden, 2005, "The reliability of inflation forecasts based on output gap estimates in real time," *Journal of Money, Credit, and Banking*, 37(3), pp. 583-601.

Orphanides, A., Porter, R.D., Reifschneider, D., Tetlow, R. and F. Finan, 2000, "Errors in the Measurement of the Output Gap and the Design of Monetary Policy," *Journal of Economics and Business*, 52(1-2), pp.117-141.

Rachel, Ł. And L.H. Summers, 2019, "On Falling Neutral Real Rates, Fiscal Policy, and the Risk of Secular Stagnation," Brookings Papers on Economic Activity, BPEA Conference Drafts, March 7-8, 2019.

Rünstler, G., 2002, "The information content of output gaps in real time: an application to the euro area," ECB Working Paper No. 182.

Timmermann, A., 2006, "An Evaluation of the World Economic Outlook Forecasts," IMF Working Paper WP/06/59.

Tosetto, E., 2008, "Revisions of Quarterly Output Gap Estimates for 15 OECD Member Countries," OECD Statistics Directorate.

Turner, D., Cavalleri, M.C., Guillemette, Y., Kopoin, A., Ollivaud, P. and Rusticelli, E., 2016, "An investigation into improving the real-time reliability of OECD output gap estimates," OECD Economics Department Working Papers, No. 1294.

Vetlov, I., Hlédik, T., Jonsson, M., Kucsera, H. and M. Pisani, 2011, "Potential Output in DSGE Models," ECB Working Paper Series No. 1351.

Appendix Figure 1. WEO Real Time Output Gap Estimates and Subsequent Revisions (Continued)

Sources: IMF *WEO* and staff calculations.
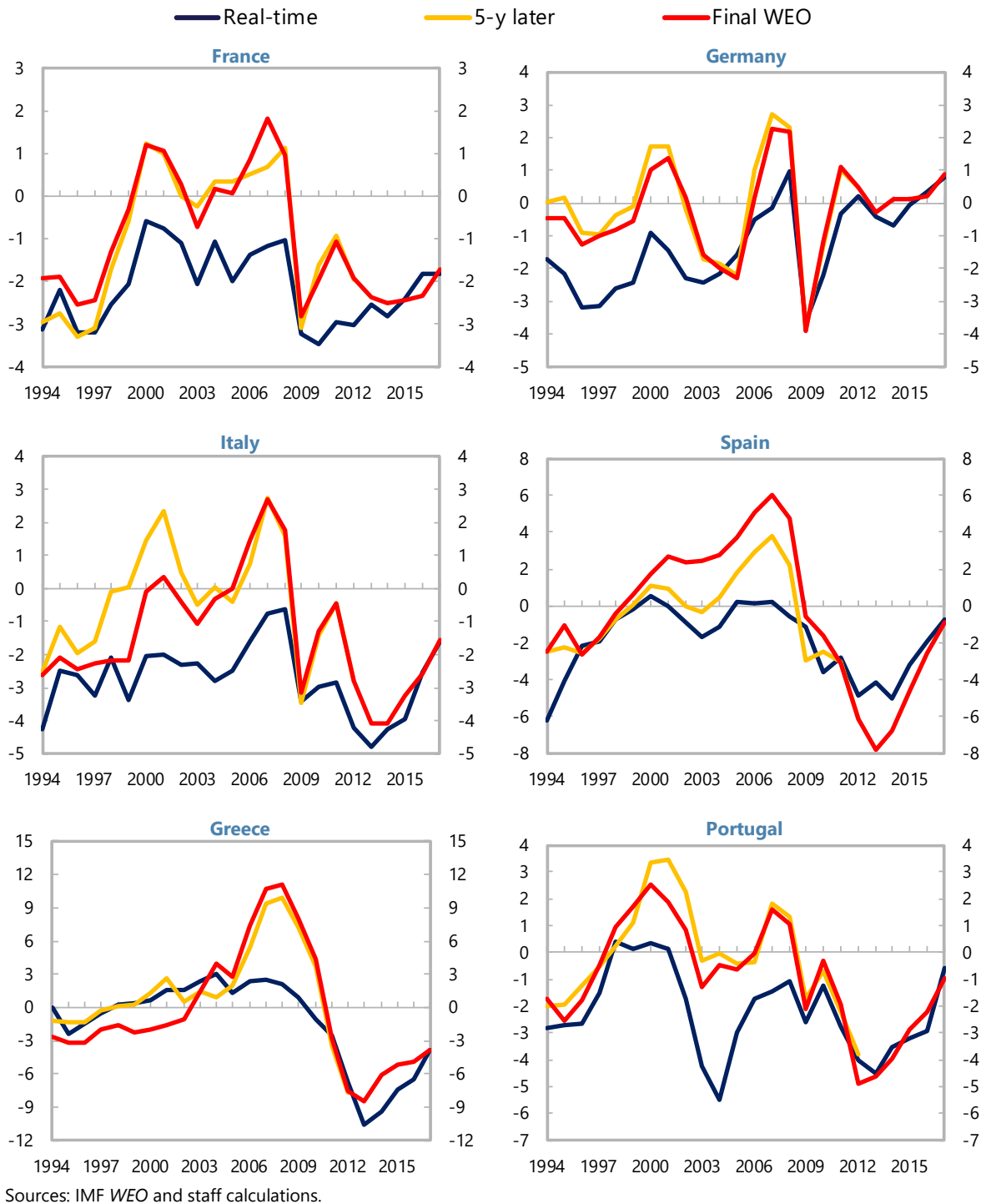
Appendix Figure 1. WEO Real Time Output Gap Estimates and Subsequent Revisions (Concluded)
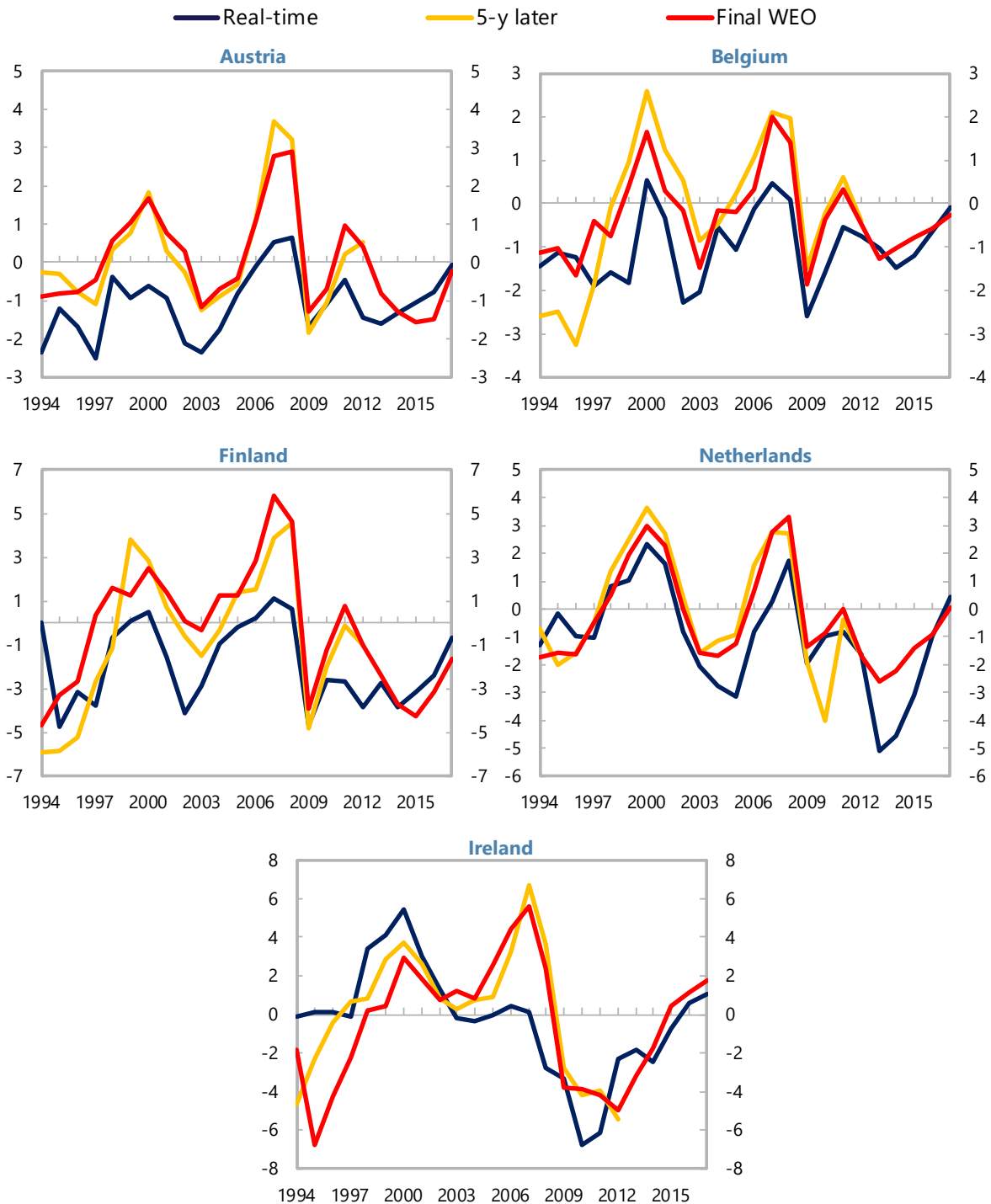
Sources: IMF *WEO* and staff calculations.

Appendix Figure 2. WEO Real Time, 2-Year Ahead and 2-Year Back Estimates
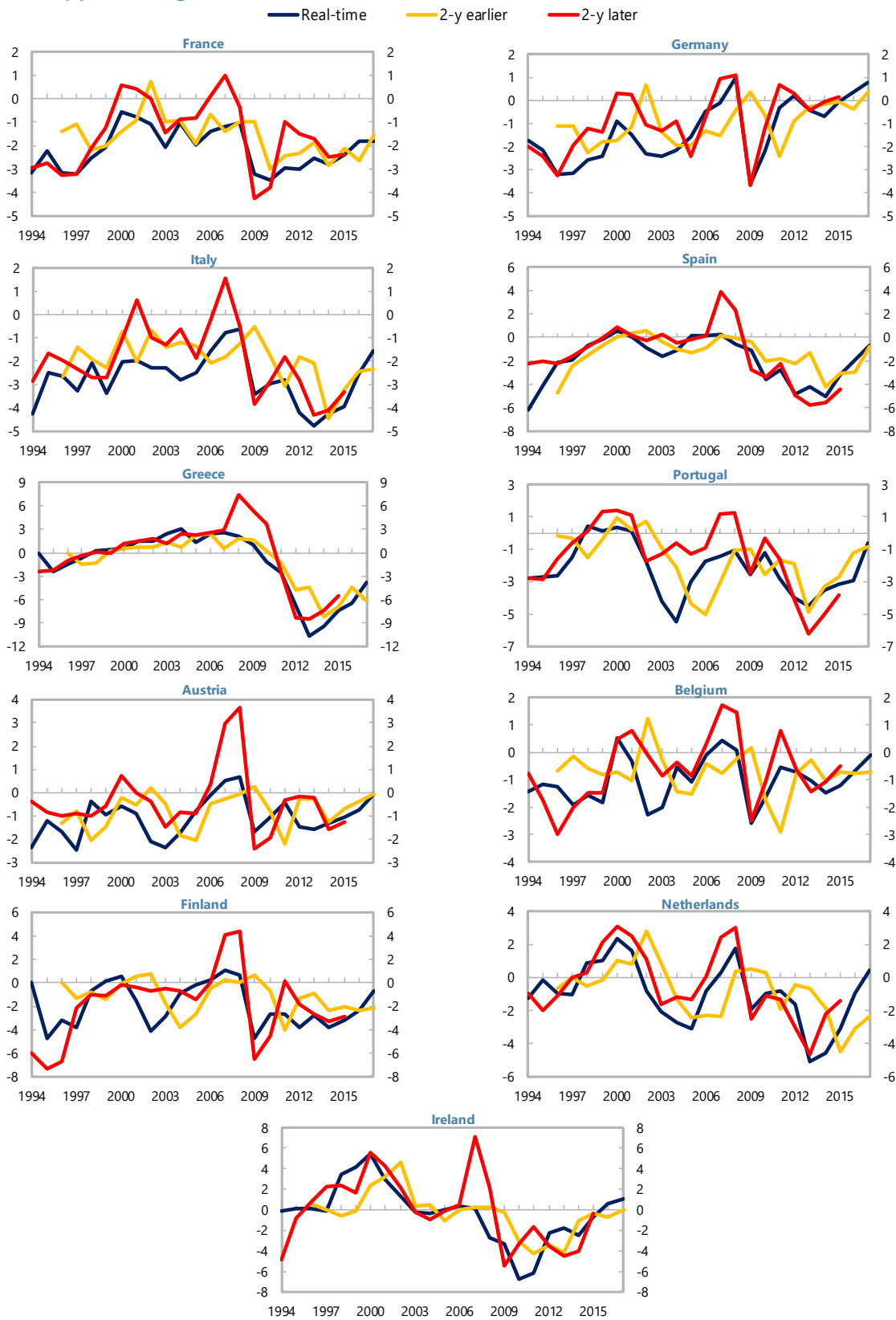
Sources: IMF *WEO* and staff calculations.

Appendix Figure 3. Real-Time and Quasi Real-Time MVF Estimates

— Real-time   — Quasi real-time
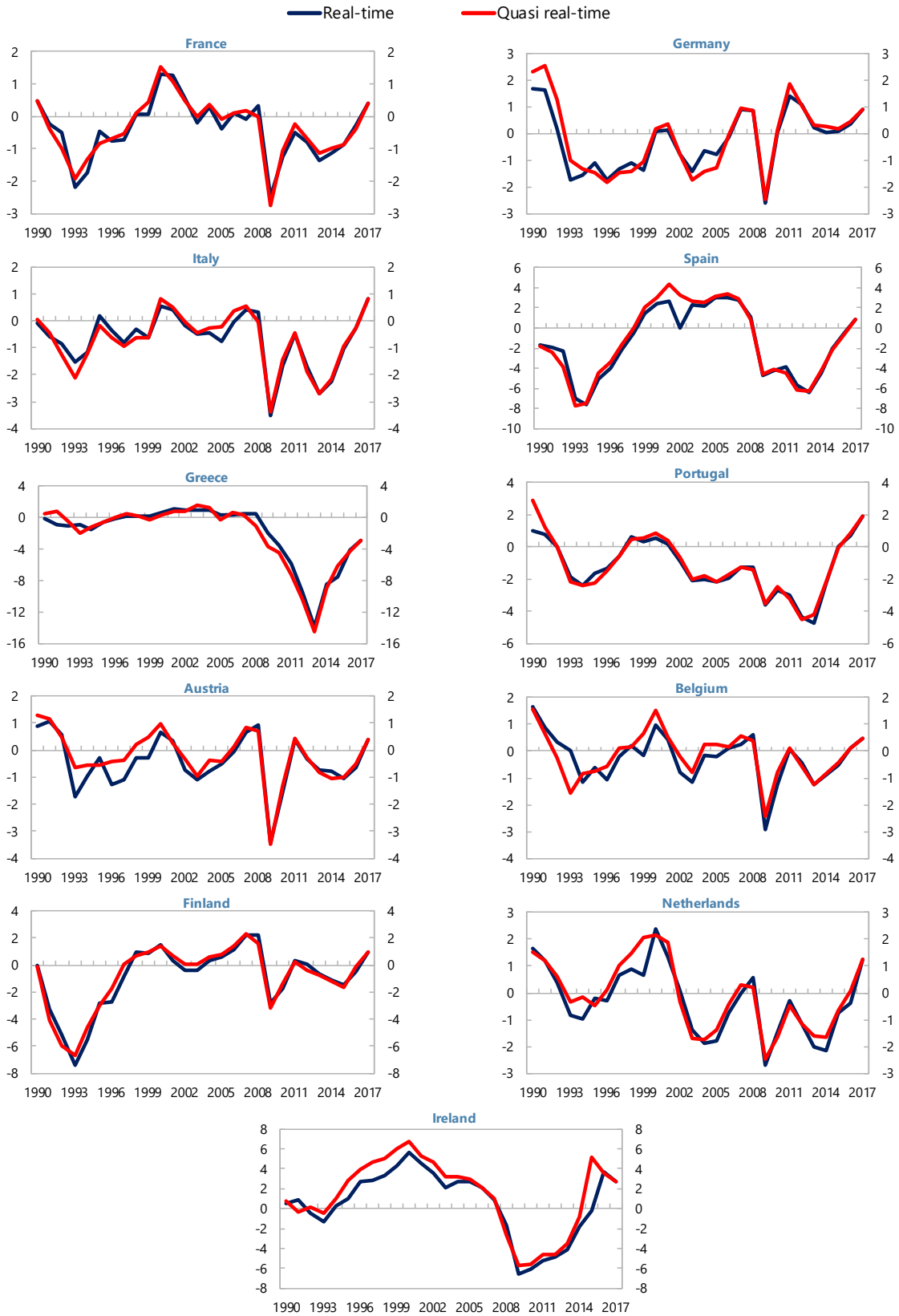
Sources: IMF *WEO* and staff calculations.

Appendix Figure 4. Vintage Forecast, Perfect Foresight, and WEO Real-Time Estimates
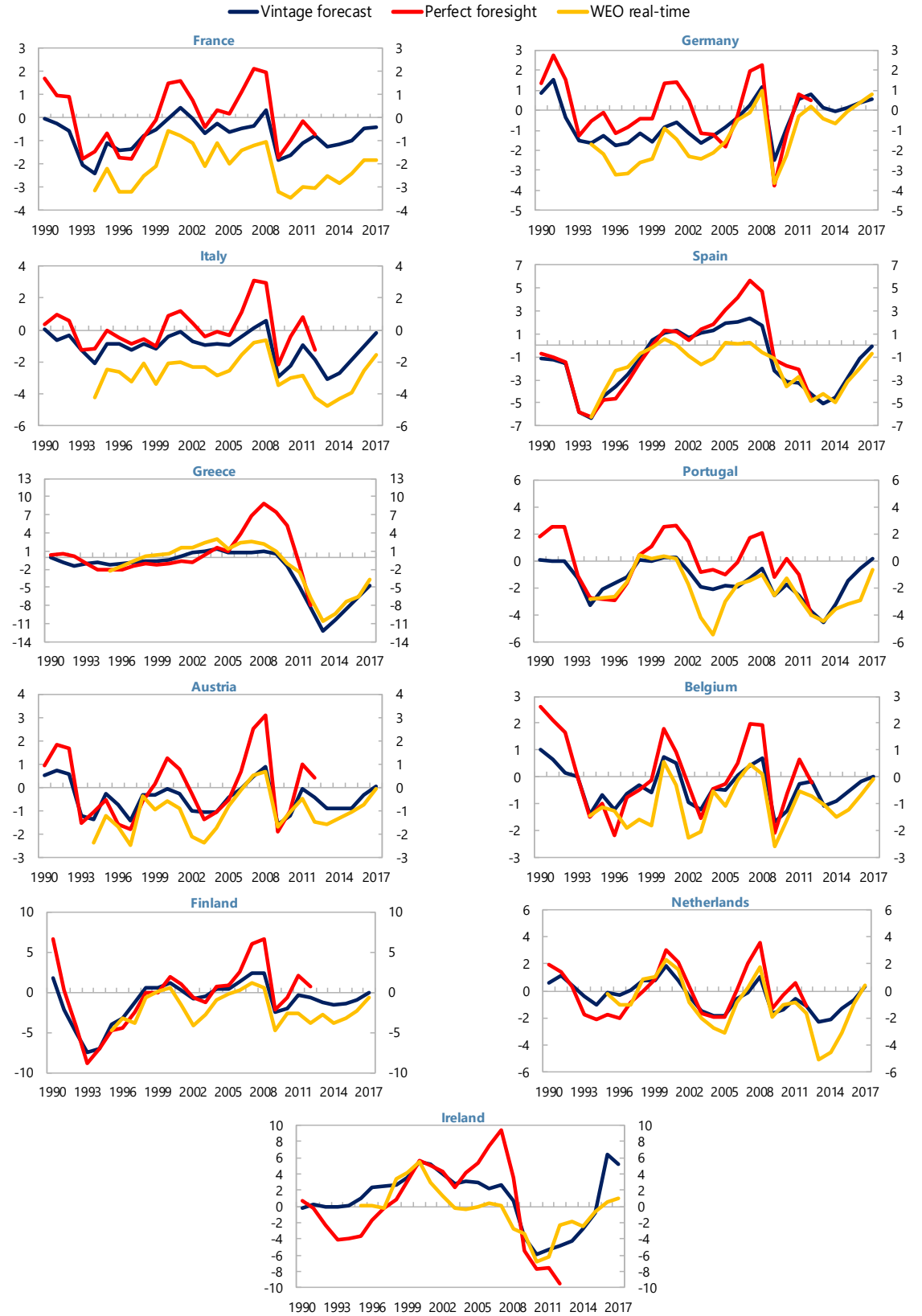
Sources: IMF *WEO* and staff calculations.

Appendix Table 1. Real-Time Output Gaps: Mean Estimates and Significance at Times of Low Real-Time Unemployment (1994–2017)

Output gap estimates when real time unemployment is below sample average (n=212)				Output gap estimates when unemployment is below NAIRU in real-time (n=80)			
	Mean	SE	p-value		Mean	SE	p-value
Euro Area	-1.1	0.27	0.00	Euro Area	-0.25		
Austria	-1.0	0.23	0.00	Austria	-1.44	0.23	0.00
Belgium	-0.8	0.21	0.00	Belgium	-0.47	0.21	0.03
France	-1.8	0.24	0.00	France			
Germany	-0.3	0.28	0.34	Germany	-0.18	0.26	0.48
Italy	-2.2	0.27	0.00	Italy	-2.81		
Luxembourg	1.5	2.03	0.45	Luxembourg	0.25	0.36	0.50
Netherlands	-0.3	0.43	0.43	Netherlands	0.45	0.43	0.30
Finland	-1.9	0.43	0.00	Finland	0.58	0.21	0.01
Greece	0.8	0.39	0.03	Greece	2.31	0.20	0.00
Ireland	1.1	0.64	0.10	Ireland	1.97	0.69	0.01
Malta	1.2	0.01	0.00	Malta			
Portugal	-1.8	0.44	0.00	Portugal	-1.45	0.48	0.00
Spain	-0.4	0.20	0.08	Spain	-0.20	0.29	0.49
Cyprus	2.5	2.89	0.39	Cyprus			
Slovakia	-1.4	0.52	0.01	Slovakia			
Estonia	-1.0	0.60	0.11	Estonia			
Latvia	-0.2	0.02	0.00	Latvia			
Lithuania	-0.7	0.64	0.30	Lithuania	-0.49	0.25	0.05
Slovenia	-0.8	0.78	0.33	Slovenia	0.29	0.91	0.76

Sources: IMF *WEO* and staff estimates.

Appendix Table 2. Potential Fiscal Implications of Negative Real Time Output Gap Bias (Percent of Potential GDP)

	Average annual bias in fiscal balances		Potential debt buildup over a sample of 19 years	
	Total MVF Bias	WEO Final Revisions	Total MVF Bias	WEO Final Revisions
Austria	0.7	0.8	13.9	15.4
Belgium	0.7	0.5	13.1	10.1
France	1.3	0.9	24.5	16.6
Germany	0.8	0.7	14.8	13.7
Italy	1.4	0.9	27.5	17.0
The Netherlands	0.4	0.4	7.7	7.2
Finland	1.3	1.4	24.8	26.9
Greece	0.1	0.5	2.1	9.3
Ireland	0.9	-0.1	16.5	-2.5
Portugal	0.8	0.8	16.0	15.5
Spain	0.9	1.2	17.9	23.4
Average	0.9	0.7	16.3	13.9

Sources: IMF *WEO* and staff estimates.
Note: Computed using the EC's semi-elasticity of budget balance to changes in the output gap (see Mourre and others, 2014).

Appendix Table 3. Real-Time Output Gaps for High and Low Debt Countries: Mean Estimates and Significance (1994–2017)

Common Sample

Real-time output gaps (unweighted)				2017 WEO vintage (unweighted)			
	Mean	p-value	DF		Mean	p-value	DF
Low debt	-0.9	0.00	155	Low debt	-0.4	0.02	155
High debt	-1.6	0.00	194	High debt	-0.5	0.02	194
Difference	-0.75	0.01	349	Difference	-0.1	0.78	349

Full Sample

Real-time output gaps (unweighted)				2017 WEO vintage (unweighted)			
	Mean	p-value	DF		Mean	p-value	DF
Low debt	-0.9	0.00	155	Low debt	-0.1	0.63	223
High debt	-1.6	0.00	194	High debt	-0.5	0.01	214
Difference	-0.75	0.01	349	Difference	-0.4	0.17	437

Sources: IMF *WEO* and staff estimates.

Appendix Table 4. Robustness of Primary Balance Regressions

	Revisions agains 2017 WEO				Revisions against (t+5) vintage			
	OLS	OLS	FE	LSDV	OLS	OLS	FE	LSDV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Primary balance (t-1)		0.595***	0.220	0.263***		0.577***	0.243	0.286***
		(0.12)	(0.16)	(0.07)		(0.12)	(0.15)	(0.07)
Output gap (final vintage)	0.655***	0.625***	0.301	0.296**				
	(0.20)	(0.16)	(0.24)	(0.15)				
Revision to real time output gap against 2017 vintage	-0.593***	-0.387**	-0.548*	-0.531***				
	(0.22)	(0.18)	(0.32)	(0.15)				
Output gap (t+5) vintage					0.718***	0.568***	0.278	0.274*
					(0.19)	(0.14)	(0.23)	(0.15)
Revision to real time output gap against (t+5) vintage					-0.507*	-0.394*	-0.547*	-0.531***
					(0.27)	(0.24)	(0.32)	(0.15)
Public debt (t-1)	0.038***	0.021***	0.100***	0.097***	0.037***	0.019***	0.092***	0.090***
	(0.01)	(0.01)	(0.03)	(0.02)	(0.01)	(0.01)	(0.03)	(0.02)
Current account (in percent of GDP)	0.426***	0.162***	0.293***	0.276***	0.423***	0.174***	0.286***	0.268***
	(0.06)	(0.05)	(0.08)	(0.08)	(0.06)	(0.05)	(0.08)	(0.08)
Government real consumption gap	0.023	-0.041	-0.017	-0.014	0.020	-0.036	-0.022	-0.019
	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)
Trade openness	-0.018*	-0.010	-0.110**	-0.103***	-0.018*	-0.011	-0.106**	-0.099***
	(0.01)	(0.01)	(0.05)	(0.03)	(0.01)	(0.01)	(0.05)	(0.03)
IMF arrangement	-1.354	0.233	-0.267	-0.231	-1.378	0.339	-0.333	-0.303
	(1.17)	(0.93)	(1.05)	(0.76)	(1.19)	(0.93)	(1.06)	(0.77)
Election year	-0.507	-0.184	-0.153	-0.156	-0.505	-0.176	-0.177	-0.178
	(0.47)	(0.34)	(0.32)	(0.33)	(0.46)	(0.34)	(0.32)	(0.33)
5Y ahead public debt forecast error	0.011	-0.018	0.034	0.032*	0.005	-0.011	0.027	0.025
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Inflation gap	0.767***	0.715***	0.615*	0.602***	0.744**	0.671**	0.685*	0.672***
	(0.28)	(0.26)	(0.33)	(0.20)	(0.30)	(0.27)	(0.36)	(0.20)
Unemployment gap	-0.335*	0.169	-0.325*	-0.305*	-0.277	0.133	-0.294*	-0.274
	(0.18)	(0.13)	(0.19)	(0.17)	(0.20)	(0.14)	(0.17)	(0.17)
NAIRU gap	-0.159	0.138	-0.483***	-0.461***	-0.143	0.115	-0.447***	-0.425***
	(0.15)	(0.14)	(0.14)	(0.13)	(0.15)	(0.14)	(0.13)	(0.13)
Fixed effects	no	no	yes	yes	no	no	yes	yes
Time effects	no	no	yes	yes	no	no	yes	yes
Constant	-0.357	-0.054	-9.188**		-0.371	0.121	-8.414**	
	(1.13)	(1.07)	(3.74)		(1.12)	(1.04)	(3.44)	
R-squared	0.49	0.69	0.84		0.50	0.68	0.84	
Adjusted R-squared	0.46	0.66	0.80		0.46	0.66	0.80	
Observations	197	195	195	195	197	195	195	195

Sources: IMF *WEO* and staff estimates.
Note: Standard errors in parentheses. ***, **, and * denote significance at 1, 5, and 10 percent, respectively. Baseline sample for 11 larger euro area countries for the period 1994–2012. The bias-corrected LSDV dynamic panel data estimates are reported with bootstrapped standard errors.