

IMF Working Paper

Strategy, Policy and Review Department

What Do We Talk About When We Talk About Output Gaps?

Prepared by Jelle Barkema, Tryggvi Gudmundsson, and Mico Mrkaic¹

Authorized for distribution by Rupa Duttagupta

November 2020

IMF Working Papers describe research in progress by the author(s) and are published to elicit comments and to encourage debate. The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

Abstract

Estimates of output gaps continue to play a key role in assessments of the stance of business cycles. This paper uses three approaches to examine the historical record of output gap measurements and their use in surveillance within the IMF. Firstly, the historical record of global output gap estimates shows a firm negative skew, in line with previous regional studies, as well as frequent historical revisions to output gap estimates. Secondly, when looking at the co-movement of output gap estimates and realized measures of slack, a positive, but limited, association is found between the two. Thirdly, text analysis techniques are deployed to assess how estimates of output gaps are used in Fund surveillance. The results reveal no strong bearing of output gap estimates on the coverage of the concept or direction of policy advice. The results suggest the need for continued caution in relying on output gaps for real-time policymaking and policy assessment.

JEL Classification Numbers: E31, E32, E47, E52, E62

Keywords: Output gaps, business cycles, monetary policy, fiscal policy

Authors' E-Mail Address: jbarkema@imf.org, Tgudmundsson@imf.org, Mmrkaic@imf.org

¹ We thank Shekhar Aiyar, Jorge Ivan Canales Kriljenko, Rupa Duttagupta, Romain Duval, and Simon Voigts for very helpful comments.

Table of Contents

Abstract	2
I. Introduction	4
II. Related Literature and Recent Debate.....	5
III. IMF Estimated Output Gaps	9
A. Summary Statistics on Output Gaps	9
B. Decomposition of GDP Gap Revisions.....	12
C. Output Gaps and Other Indications of Slack.....	14
IV. Output Gap Coverage in IMF Bilateral Surveillance	20
V. Concluding Thoughts	33
References	36

I. INTRODUCTION

The macroeconomic environment after the Global Financial Crisis (GFC) has led to a rethink of policymakers' ability to manage the business cycle. Low and declining rates of unemployment have not been accompanied with any noticeable rise in inflation, particularly in advanced economies, while continued monetary policy accommodation and a large increase in central banks' balance sheets have been unable to reinvigorate growth. These puzzles have led to questions about the framework for estimating economic slack, or resource utilization compared to its potential, in the economy.

The ongoing Covid-19 crisis has further exacerbated these issues with uncertainty about the likely path of growth very high. Difficulties in measuring the relative falls in demand and supply, and the unusual nature of the pandemic-induced downturn, further complicate estimates of slack with dispersions in growth forecasts subsequently very high. The unknown extent to which the downturn entails temporary or permanent damage to activity also muddles the task of deciding the required amount of policy support by authorities.

The most common conceptual framework for slack is that of the output gap, which relates an economy's potential output, on the one hand, to its actual output, on the other, and is defined as the difference between the two. A positive gap represents an overheating economy, whereby resource utilization is above its steady state feasibility, whereas a negative gap signifies underused resources.

To the extent that the output gap is used as an indicator of slack, it plays a crucial role in guiding the optimal policy path at any point in time. An underestimated output gap will, for example, result in a policy setting that is too loose, other things equal. This will in turn result in overheating pressures, resource misallocation, and a subsequent need to tighten policy to a greater extent than previously planned. The ensuing rise in macroeconomic volatility would have been avoided with a more precise assessment of potential output. Similarly, an overestimated output gap that results in a tighter than required policy stance would lead to resource underutilization, unnecessary unemployment, and lower output than aggregate demand and supply conditions in the economy would have tolerated.

After a long period of estimated negative output gaps since the global financial crisis, there was a growing divide in policy circles on the state of global business cycles (see Section II). Most major economies were estimated by the IMF and others (see for example the October 2019 *World Economic Outlook*) to have been operating at or slightly above potential by the end of 2019. This assessment, induced in part by persistently high levels of capacity utilization and appreciably low unemployment rates, would indicate that further accommodation of fiscal and monetary policy would be counterproductive, potentially resulting in overheating pressures. In contrast, this view was challenged by others (see e.g. Brooks and Basile (2019)) who argued that the absence of inflationary pressures indicated

that output gaps were still negative, thus warranting continued emphasis on policy support to raise demand and output up to its potential².

Against the backdrop of these debates, the purposes of this paper are more modest. Our aim is to review the historical distribution of output gaps based on IMF surveillance, assess the degree of reliance on this measure as the indicator of slack, and how output gap estimates have compared to other measures of over- or underheating. These tasks help assess the extent to which real-time policymaking can use output gaps as a reliable guide for conducting policy. Our findings suggest that the distribution of output gap estimates globally contains a negative skew, that there is a positive but imperfect relationship between output gaps and other measures of slack, and that output gaps discussions are widespread in IMF surveillance.

Before proceeding, the caveat that business cycle estimates are only one input for policymaking should be stressed. Policy recommendations may rightly ignore any signal coming from output gap estimates if other concerns, such as debt sustainability, are thought to be of greater importance at that point. This implies that the output gap is not the sole arbiter of any policy stance. All of these caveats are well known by both supporters and critics of the output gap as a gauge of slack. We avoid this debate and instead focus on the role of the output gap in informing policy advice in IMF surveillance.

The rest of the paper is organized as follows. Section II discusses the recent debate and the surrounding literature on the viability of output gaps as measures of slack. Section III provides summary statistics on IMF-estimated global output gaps, including by region and income level. Section IV assesses how closely output gap estimates coincide with other more direct measures of slack with simple empirical exercises. Section V analyzes output gap coverage in IMF surveillance using text analysis. Section VI concludes.

II. RELATED LITERATURE AND RECENT DEBATE

The pre-Covid-19 assessment of output levels being generally close to or above potential in most advanced economies invigorated the debate about the output gap concept with disagreements on the amount of slack and the consequences for policy (see e.g. IIF(X)). However, such discussions are certainly not new and there is a rich tradition on the topic in the literature going back decades³. Okun (1962) provided an argument for the use of potential output, covering general aspects such as measurement and statistical estimates. Pesek (1963) discusses the debate in the US at the time, which received a boost when the Council of

² This argument has gained ground recently amongst a wider audience. In laying out considerations for the ECB's ongoing monetary policy strategy review, Lagarde (2020) cites potential mismeasurement of slack as a key possibility behind the weak response of inflation to macroeconomic improvements in the run up to the Covid-19 crisis. Within this argument, Lagarde discusses "hidden slack" and "revisions to potential output which mistook cyclical changes for structural trends" among the hypotheses.

³ In a critical piece, Plosser and Schwert (1979) claim that potential output had "played a central role in discussions and the implementation of economic policy for at least fifteen years". Their pushback against the use of potential output include criticisms that are still invoked today, including that "most efforts to estimate potential output [...] are essentially equivalent to trend extrapolation of output" (p. 185).

Economic Advisers started to use potential output in its analysis, and argues that "a concept which is so frequently used and so readily understood should be worth understanding"⁴. Kuh (1966) subsequently covered some of the early methodological issues surrounding potential output and the output gap, but concluded that the empirical bases of these parameters "are shaky enough so that excessive confidence ought not to be placed on the present set of point estimates which, however, seem solidly enough based to illustrate an appropriate set of procedures"⁵.

More recently, the literature has added greater sophistication to the same basic idea of a potential maximum level of output. However, the same reservations remain, namely that point estimates do not provide a reliable guide and that estimates of potential output can vary wildly, particularly in real time. Despite these reservations, the prevalence of Taylor rules and other optimal policy measures have further embedded output gap measurements into policymaking in recent decades with the concept used as a barometer by which the relative tightness of policy, monetary policy, is measured (see e.g. Mertens and Williams (2020)). Even in 2020, in the aftermath of Covid-19, the extent to which the crisis represents a greater supply or demand shock has been widely debated, not least due to the view that should the demand hit be larger, a negative output gap opens up with implications for the stance of policy⁶.

Despite some methodological progress, the framework for estimating the output gap has not changed radically in recent years. The common method of using a production function approach dates back several decades, while the other most common approach of the Hodrick-Prescott (HP) filter originated in research from the early 1980s and gained popularity in the following years⁷. Some studies have found signs of predictive power using simple output gaps. Claus (2000) looks at the case of whether the output gap in New Zealand is a useful indicator of inflation. Using simple reduced form models, he finds that the gap does provide a signal of future inflation. More recently, however, Jarocinski and Lenza (2018) attempt to reverse-engineer the question in a way by computing how big of an output gap is needed for the euro area to account for realized inflation. Using a Bayesian dynamic factor model, they find that the output gap that can plausibly coincide with the weak inflation behavior is significantly larger than traditional estimates⁸.

⁴ Pesek (1963) p. 297.

⁵ Kuh (1966) p. 773.

⁶ See Bekaert et al (2020) for a discussion and estimation of the relative demand and supply shocks involved during the Covid-19 crisis.

⁷ See Congdon (2008) for a history of the output gap concept. Kydland and Prescott (1982) and Hodrick and Prescott (1997) provide much of the impetus for later use of the filters, although the general concept of filtering goes significantly further back.

⁸ Using several specifications, they find that the model with best out-of-sample fit produced an approximate gap of -6% in 2014-2015 compared to common estimates in the range of -2-3%.

Such skepticism is common and there is a long history documenting potential difficulties with the concept and its shortcomings. Orphanides and van Norden (2002) document that revisions to the estimated gap are of the same order of magnitude as the gap itself, and that despite a myriad of attempts to improve the estimation of gaps, multivariate methods are no more reliable than univariate ones⁹. They also highlight the particularly challenging issue of identifying turning points in the business cycle and find that all methods of estimation under consideration underestimate the output gap at cyclical peaks.

These various criticisms have led many to emphasize the inherent uncertainty in slack and output gap estimates. Gordon (1997) looks at the closely related concept of NAIRU for the US and argues that it remains a helpful variable despite the large uncertainty in its estimation. Smets (2002) looks at output gap measurement errors and their effect on monetary policy rules. He finds that output gap uncertainty can have a significant effect on optimal monetary policy responses.

On the methodological front, Hamilton (2018) presents a forceful critique of the HP filter. In another line of criticism, Coibon et al (2018) argue that real-time estimates of potential output respond similarly to transitory and permanent shocks. Revisions in estimates therefore provided little signal on the permanence of output losses. While they acknowledge the multiple constraints in real-time estimates, they argue that attempts by institutions such as the IMF to strip out cyclical variation in output and identify long-run changes have been “largely unsuccessful”¹⁰.

Yet another line of work notes the asymmetric history of output gap estimates. Looking at the record for European countries, Kangur et al (2019) find that there is a negative skew in estimates, meaning that countries are more often thought to be operating below potential than above it¹¹. A symmetric business cycle would presumably not produce these results although some theories, including the plucking model, are consistent with such behavior¹². Kangur et al also find that output gap estimates have limited predictive power for inflation. Separately, Cerra and Saxena (2017) argue that downturns tend to inflict permanent damage to trend output, thus rendering output gaps “extremely difficult to measure and more difficult to interpret”¹³.

⁹ In a follow-up paper, Orphanides and van Norden (2005) cast further doubt on the predictive power of output gaps for inflation, particularly in real time.

¹⁰ Coibon et al (2018) p. 405.

¹¹ Grigoli et al (2015) produce similar findings for a broader set of countries.

¹² See for example Dupraz et al (forthcoming) wherein economic fluctuations represent drops below the economy’s full potential ceiling, as opposed to fluctuations above and below the potential. Aiyar and Voigts (2019) motivate a negative mean output gap as a result of downward nominal wage rigidity.

¹³ Cerra and Saxena (2017) p. 12.

Due to these, and other, critiques of the traditional approaches to output gaps, several studies have suggested alternative methods of measuring output gaps. One of the most common ones in the recent literature is that of multivariate filters, which expand on the HP filter by incorporating numerous other variables. Benes et al (2010) and Blagrove et al (2015) represent attempts to improve estimates using multivariate filters. Borio et al (2013) and later Borio et al (2016) argue for the use of information about the financial cycle in measures of potential output. Berger et al (2015) also focus on the role of financial variables. Rabanal and Sanjani (2015) depart from the filter approach by estimating a DSGE model for the euro area to analyze financial shocks. In a recent study, Banbura and Bobeica (2020) find that historical estimates of slack by international economic institutions, namely the EC, the ECB, the IMF, and the OECD, improve the fit compared to simple filter-based approaches¹⁴.

Despite these attempts, however, there is no consensus on a superior approach that addresses in full the shortcomings of the traditional approaches. Chen and Gornicka (2020) perform a horse race and find that a variant of a structural VAR model provides relatively smaller out-of-sample forecast errors. They too, however, urge caution in basing policy decisions on point estimates of output gaps.

The recent policy debate, as opposed to the methodological one, in turn revolves around the current growth environment and the assessment of current potential growth. The debate was ignited in particular by Brooks and Basile (2019) who argued that estimates of closed or closing European output gaps from institutions including the IMF, the European Commission, and the OECD, “make little sense” in light of the divergence of growth performance on the continent and the lack of inflation. According to Brooks and Basile, potential output estimates do not capture the level of output consistent with stable inflation but rather just the realized recent output. The authors conclude that considerable slack remained in peripheral Europe as of end 2019.

An important subsequent policy implication could be that depressed potential output becomes self-fulfilling, as weak realized growth feeds into estimates of potential. Tight monetary and fiscal policy would thus ensure a reduction in potential growth.¹⁵ In response, Buti et al (2019) acknowledge issues and challenges surrounding potential output measurement and the inherent uncertainty involved. They also highlight the imperfect relationship between output and inflation and argue that a wide difference in potential output between countries is a perfectly plausible possibility¹⁶.

¹⁴ Section 3 of this paper provides more detailed information on the IMF estimates.

¹⁵ See also Benigno and Fornaro (2017) who present a traditional framework where pessimistic expectations can lead to permanent slumps.

¹⁶ Coibon, Gorodnichenko and Ulate (2018) provides a similar discussion for the US case. They argue that methods used by key agencies, such as the IMF, the CBO, and the Federal Reserve, mix up supply and demand shocks. They conclude, similarly to Brooks and Basile (2019) that output gap estimates at the time were too low and that greater accommodation was required by policymakers which would in turn boost household incomes significantly.

Our goal is not to relitigate the methodological debate but rather take a relatively general look at the historical evidence, developments of output gap estimates, and the context in which output gaps are discussed and how they influence policy discussions. The next section documents the distribution of output gap estimates from IMF surveillance from the available data, before moving on to analysis of the relationship of output gaps to other measures of overheating as well as coverage of output gaps in country surveillance.

III. IMF ESTIMATED OUTPUT GAPS

A. Summary Statistics on Output Gaps

As part of its bilateral surveillance function, the IMF produces forecasts of output gaps which inform part of the policy recommendations in areas including fiscal and monetary policy. In this section, we look at the historical evidence of output gaps on a global basis according to IMF estimates and how that assessment has evolved over time and across regions¹⁷. Country teams have discretion in terms of methodology when estimating the output gap. The estimates have included filter-based approaches, both univariate and multivariate, production function methods, as well as a variety of other modeling approaches and the application of judgement.

We also look at revisions to output gap estimates. Revisions can play a crucial role as policy advice is formulated in real time and subsequent revisions to estimated gaps could render the initial policy advice less reliable. The output gap is a crucial component in the formulation of this policy advice. As noted by the IMF's Independent Evaluation Office (IEO) in 2014, it is "a key indicator of the degree of slack in the economy and is typically used in short-term forecasts of inflation and the measurement of cyclically adjusted fiscal and current account balances, factors that are critical in the IMF's policy advice to member countries."¹⁸

Table 1 below shows the aggregate estimates of the real-time output gaps using annual data for all available countries in the IMF's *World Economic Outlook* (WEO) database. The phrase "real-time" refers to the gaps as they were reported at each point in time, for example the output gap for any given country in 2007 is the one reported in 2007.¹⁹ Subsequent revision to historical estimates mean that the latest output gap, for example the gap for 2007 reported in 2018, may differ significantly from the real-time estimate. The "final revision" column thus represents the latest available estimate for any given year. This inevitably means that the most recent data may yet be revised whereas the older estimates are unlikely to undergo significant future revisions.

¹⁷ Kangur et al (2019) provide a recent overview of the IMF-estimated output gaps for Europe.

¹⁸ IEO (2014) p. 28.

¹⁹ As in other papers, we make the distinction between real time estimates and final or revised estimates. The former implies an estimate at time t of the output gap at time t whereas the latter is the time $t+n$ estimate of the output gap at time t .

Table 1. GDP Gap Estimates, Summary Statistics by Region /1

	Statistic	Real time	One year later	Final estimate
EUR	Mean	-1.07	-0.79	-0.24
	SE mean	0.10	0.10	0.10
	10th pctl	-3.66	-3.40	-3.31
	Median	-0.86	-0.70	-0.25
	90th pctl	1.04	1.51	2.92
MCD	Mean	0.01	0.74	2.71
	SE mean	0.44	0.51	0.77
	10th pctl	-2.53	-1.82	-2.74
	Median	-0.73	-0.42	0.04
	90th pctl	5.20	6.43	8.61
AFR	Mean	-1.05	-0.40	0.16
	SE mean	0.40	0.46	0.15
	10th pctl	-2.61	-1.23	-1.66
	Median	-0.46	-0.32	0.00
	90th pctl	0.55	2.27	2.32
WHD	Mean	-1.04	-0.68	-0.13
	SE mean	0.13	0.13	0.13
	10th pctl	-3.29	-3.03	-3.22
	Median	-0.63	-0.44	-0.10
	90th pctl	0.86	1.28	3.24
APD	Mean	-1.01	-0.74	0.05
	SE mean	0.13	0.13	0.17
	10th pctl	-4.15	-3.30	-2.90
	Median	-0.42	-0.35	-0.09
	90th pctl	0.81	1.15	2.73
All	Mean	-1.02	-0.69	0.11
	SE mean	0.07	0.07	0.09
	10th pctl	-3.53	-3.30	-3.12
	Median	-0.70	-0.53	-0.08
	90th pctl	0.95	1.43	3.19

/1 Sample 1995-2018

Our original sample includes 197 countries and territories covering the sample period of 1995-2018. This represents all countries for which real-time estimates are available during the sample period and excludes 2019 due to the recency of the data. Some countries have longer data series available than others, introducing the possibility for some bias to overcome so we also look at different sub-samples and various slices of the data. In total, our data includes 1,119 annual estimates or an average of just over six per country.

A few facts are notable from these numbers. Looking first to the real time estimates, the distribution is clearly not centered around zero, with the mean and median both firmly in

negative territory. For our total sample, the median output gap is -0.7% for the 1995-2018 period, suggesting that the world economy has been assessed to have operated with a certain degree of slack over time²⁰. Looking at the different regions, this negative skew persists, apart from countries in the Middle East and Central Asia where the mean is approximately zero, although the median is still firmly negative which in turn suggests the influence of outliers. This pattern does not change even when the years of major economic and financial crises are excluded from the data (e.g., 1997-99, 2000-02, 2008-10).

Turning to revisions, we see that there is a clear tendency for output gap estimates to shift towards zero over time. This is the case for both $t+1$ estimates, i.e. those made one year after the initial estimates, as well as the final revisions. In the case of the latter, output gaps have broadly reached zero by the time of final estimates.

The finding of negative mean and median output gaps is in line with previous estimates for specific regions (see Section II for a discussion of prior findings). However, the result could be driven both by a sharp drop in actual output and estimates of potential output during the global recession of 2008-2009.²¹ Indeed, estimates are higher in the pre-crisis period, but they still retain a negative skew in both periods.

Yet another slicing of the data looks at estimates by income level, instead of geographic region. Table 2 shows that there is not much difference between advanced economies, on the one hand, and emerging markets and low-income countries, on the other²². Both groups retain the familiar property of negative means and upward revisions. Furthermore, a similar breakdown between pre-crisis and post-crisis periods yields the same results as by geographic distribution, namely that the negative skew is higher pre-crisis but present in both periods.

²⁰ Kangur et al (2019) find a mean real-time output gap of -1.7% for the euro area between 1994 and 2017.

²¹ There is a relative lack of available data for Africa and the Middle East, compared to other regions. For the whole sample period there are only 166 final output gap estimates for countries in the African department and 43 real-time estimates.

²² In Section V we also look at output gap coverage by income group.

Table 2. GDP Gap Estimates, Summary Statistics by Income Level /1

	Statistic	Real time	One year later	Final estimate
AE	Mean	-1.06	-0.77	-0.19
	SE mean	0.09	0.09	0.10
	10th pctl	-3.63	-3.38	-3.07
	Median	-0.81	-0.63	-0.20
	90th pctl	1.02	1.40	2.59
EM & LIC	Mean	-1.02	-0.69	0.11
	SE mean	0.07	0.07	0.09
	10th pctl	-3.53	-3.30	-3.12
	Median	-0.70	-0.53	-0.08
	90th pctl	0.95	1.43	3.19
All	Mean	-1.02	-0.69	0.11
	SE mean	0.07	0.07	0.09
	10th pctl	-3.53	-3.30	-3.12
	Median	-0.70	-0.53	-0.08
	90th pctl	0.95	1.43	3.19

/1 Sample 1995-2018

B. Decomposition of GDP Gap Revisions

We next decompose output gap revisions into the revisions of potential and actual GDP respectively to gain further insight into the nature of overall revisions. Table 3 below shows the extent of changes to initial output gaps over time. The table separates the total amount of revisions to changes in actual underlying data, on the one hand, and updated estimates in potential GDP, which affect the assessment of the output gap, on the other. We see that, firstly, the revisions are large in absolute size and, secondly, the bulk of the revisions is due to updated assessments of potential GDP downwards. This is also broadly the case when looked at by income level and region. The relationship also holds when we look at the full sample and post-GFC subsample, in Tables 4 and 5²³.

Table 3. Contributions of GDP revisions to the revision of GDP gap /1

Region	Next year	Final
EUR	0.58	0.54
MCD	0.27	0.21
AFR	0.29	0.01
WHD	0.21	0.25
APD	0.33	0.49
All	0.11	0.36

/1 Sample 1995-2018

²³ The contributions are calculated using the RES WEO database of GDP revisions.

Table 4. Contributions of GDP revisions to the revision of GDP gap /1

Region	Next year	Final
EUR	1.00	0.07
MCD	0.01	0.12
AFR	0.00	0.20
WHD	0.13	0.00
APD	1.55	1.43
All	0.27	0.36

/1 Sample 2009-18

Table 5. Contributions of GDP revisions to the revision of GDP gap /1

Region	Next year	Final
AE	0.02	0.35
EM & LIC	0.53	0.32
All	0.11	0.36

/1 Sample 1995-2018

An important caveat to note at this stage is the potential endogeneity of potential output and actual output. To the extent that a lack of growth reduces long-term growth, for example via hysteresis effects, it is to be expected that a shift in actual growth affects an economy's potential. Similarly, if a short-term boost in demand, for example via productive investment, involves long term gains, a positive relationship between potential and actual output is reasonable. The traditional approach of attributing potential growth to structural supply-side factors would diminish this channel but recently there has been a greater emphasis for the possibility of demand affecting supply, or a de facto partial inverse of Say's Law²⁴.

In terms of variation of output gap estimates, there are significant commonalities across major institutions in estimating output gaps, both in terms of methodology and results. The findings above are thus not an institution-specific result²⁵. This finding of substantial revisions is also in line with those found for specific subregions and highlights further the difficulty of basing policy advice or implementation on real-time estimates of output gaps (see discussion in previous section)²⁶.

²⁴ See for example Rachel and Summers (2019) or Eggertsson et al (2019) for discussions of persistence in demand shortfalls and accompanying untraditional output dynamics.

²⁵ Tereanu et al (2014) find that the European Commission's output gap revisions amounted to 1.3% of potential GDP on average during the period of 2003 to 2012. Coibion et al (2018) look at estimates from a wide set of institutions. Turner et al (2016) discuss the large revisions to OECD-estimated output gaps and attempts to improve their real-time reliability within a production function approach.

²⁶ As discussed in the previous section, Kangur et al (2019) find a negative skew and significant revisions for European countries and Grigoli et al (2015) for a broader set of countries.

The general finding of negative output gaps with frequent subsequent revisions adds tentative support to the various criticisms discussed in Section II. The finding of average negative output gaps, coupled with the common view that policy be loosened during periods of negative output gaps and tightened when confronted with positive output gaps would, other things equal, lead to more frequent easing periods than tightening. However, it should be kept in mind that there are at times non-cyclical reasons for adjustments to policy. This would include debt sustainability concerns and unanchored inflation. One would therefore expect an imperfect relationship here. The size and frequency of revisions, in turn, suggest that, at best, estimates only become reliable indicators of slack long after the period in question. Whether estimates of output gaps are indeed credible indicators of slack also depends on the extent of co-movement between the estimates and other measurable business cycle indicators. We turn to this question in the next section.

C. Output Gaps and Other Indications of Slack

The output gap is an unobserved variable, even in hindsight. It is the difference between an economy's actual output, which is observed, and its potential output, which must be estimated and can be refined but is not actually observed. As such, the usefulness of output gap as a measure of slack can be assessed by comparing it to other indicators of under- or overheating.

In this section we look at the extent to which output gap estimates comove with other indicators of interest that are observable²⁷. For this, we focus on inflation, unemployment, and current account balances²⁸. These three variables share several appealing characteristics. Most crucially, they provide an indication of possible under- or overheating of the economy, they are in and of themselves variables that are of interest to policymakers and are available across a large sample of countries and over a long period of time.

There are inevitable complicating factors when looking at these variables as indicators of slack. A clear example is labor market rigidities which vary across countries. A high level of unemployment coinciding with a negative output gap may therefore simply indicate that the non-accelerating inflation rate of unemployment (NAIRU) in that country is high. Similarly, with inflation, a closed output gap may correspond to different levels of inflation among countries. We attempt to address such concerns by also looking at changes in the variables. Despite some exceptions, a *rise* in inflation would *ceteris paribus* be expected to go along with a *higher* output gap.

The variables in question are also determined by factors going beyond the level of slack. Inflation may, for example, rise or fall depending on exchange rate developments which are unrelated to the business cycle, multinationals will affect the current account data, and

²⁷ Hristov et al (2017) perform a related analysis for a sample of European countries. In line with many other studies in the area, they highlight the inherent uncertainty in output gap estimation.

²⁸ Capacity utilization would be another candidate but availability is sparse for our sample of countries.

unemployment responds to structural developments in labor markets. As for the attractiveness of the individual variables, the current account balance is perhaps the least direct measure of slack as it correlates only partially to the economic cycle. However, we include it since overheating can manifest itself through external balances including trade and exchange rates, leading to seemingly balanced developments of domestic variables such as inflation or unemployment.

These simple empirical exercises are not intended to determine causality between the variables at hand. Where discretion is used to estimate the output gap, it is itself in part based on other indicators of slack so the gap can partially be thought of as a function of the slack measures. Similar to the Philips curve, the purpose of this exercise is simply to convey the extent to which there is co-movement of the variables.

It should also be noted that a finding of limited co-movement between output gaps and measures such as inflation does not automatically signal a mis-specified output gap.²⁹ This would, for example, be the case for emerging markets that tend to experience exchange rate appreciation pressures in times of expansion. This appreciation would in turn dampen inflation and in effect “hide” the underlying pressures on capacity. While the stronger currency would in theory lead to a decline in the country’s current account balance, which is another of our indicators, that relationship is also imperfect.

Another practical consideration is the flattening of the Phillips curve (see e.g. Hooper et al (2019)). If the relationship between real variables, such as output and employment, and prices has become weaker or exhibits non-linear characteristics, a fairly weak link should be expected when comparing output gaps and measures of overheating, at least compared to the previous period which exhibited a steeper Phillips curve. Despite these various caveats, one would a priori expect a positive, yet imperfect, co-movement of output gaps and other indicators of slack over time.

Table 6 below shows summary statistics for our universe of *World Economic Outlook* output gaps from the previous section and our three observable variables. We use the real time output gap estimates, as opposed to revised, to assess the estimate of slack as it happened as opposed to in hindsight. To reduce the influence of outliers we calculate the Spearman correlation of ranks in addition to the traditional Pearson correlation. We initially focus on the real-time gaps before turning our attention to the “final” gaps.

²⁹ In principle, the co-movement of the output gap and other measures of slack could be due to the fact that the gap is estimated using such other measures, resulting in the problem of identification. However, in practice, output gaps are typically estimated using the HP filter, which does not use any other measures of slack.

Table 6. Correlations of Measures of Slack, 2009-18 (* is significant at 5% value)

	Spearman		Pearson	
	<i>Real time</i>	<i>Final</i>	<i>Real time</i>	<i>Final</i>
GDP gap				
Inflation	0.1266*	0.1966**	0.0042	-0.0033
CA balance	0.0723	-0.0063*	0.1077*	0.0447
Unemployment rate	-0.3967*	-0.2829*	-0.3434*	-0.2537*

We see that for the total sample the correlation between the gaps on the one hand and the directly observable measures of overheating on the other is limited, but broadly consistent with economic theory.³⁰ The real time output gaps are significantly correlated with unemployment (negatively) and, once corrected for outliers, also with inflation.³¹ An economy operating at a higher capacity, relative to its potential, tends to experience higher inflation and lower unemployment.

To determine whether extensive revisions, documented in the previous sections, result in a more visible relationship between output gaps and our three variables, we also estimate the correlations using the “final” output gap estimates. Instead of the real-time estimates, we therefore use the latest estimate for historical output gaps. This does, however, have the drawback that the most recent data can be expected to be revised in the future whereas the older data is likely final at this point. Nonetheless, the data indicate whether the passage of time and new information align the latest estimates of output gaps more with the observable variables.

As we can see in Table 7 below, the co-movement is improved somewhat by using final estimates. Compared to the real-time estimates, the link between higher output gaps and tighter labor markets has weakened while the other variables show better agreement with the expected prior values. Furthermore, as mentioned above, year-over-year changes in the output gap on the one hand and the three variables on the other, as opposed to level comparisons, may provide more information on the state of the business cycle. The table shows the same relationship using first differences. We do indeed see evidence of stronger co-movement. This suggests that it is the change in slack that output gaps can help identify, rather than the outright level at any given point in time.

Table 7. Correlations of Measures of Slack, differences, 2009-18

	Spearman		Pearson	
	<i>Real time</i>	<i>Final</i>	<i>Real time</i>	<i>Final</i>
GDP gap				
Inflation	0.1560*	0.2550*	0.0123	0.0123
CA balance	-0.1355*	-0.1553*	-0.1082*	-0.0879*
Unemployment rate	-0.4064*	-0.5174*	-0.3283*	-0.4833

³⁰ Note that the decade-averages do not average out to the total sample numbers as this is an unbalanced panel. For example, we have 152 observations for the 1991-2000 period but 735 for 2011-2018, leading to a greater total contribution from the latter years.

³¹ We use Spearman rank correlation to reduce the impact of outliers.

Before moving to geographic variation, Table 8 below looks at the development of the correlations over time. We look again at both Pearson and Spearman correlations, using both real time and final estimates of the output gap. We see no great trend towards increased or decreased co-movement for any of the measures, nor do we see trends in significance. The Spearman results for final gaps, which depend less on outliers and focus on what is deemed to be in hindsight a more accurate assessment of the output gap, does however, show general significance and the expected sign for most subperiods and indicators.

Table 8a. Pearson correlation with real time GDP gap

	1995-99	2000-04	2005-09	2010-14	2015-18
Inflation	-0.0019	0.2802*	0.0139	0.1821*	-0.0087
CA	-0.0029	-0.1936*	0.0418	0.0876	0.1688*
Unemployment	-0.2298*	-0.0895	-0.3283*	-0.4414*	-0.3055*

Table 8b. Spearman correlation with real time GDP gap

	1995-99	2000-04	2005-09	2010-14	2015-18
Inflation	0.1996*	0.4328*	0.2218*	0.2610*	0.1142*
CA	-0.2085*	-0.2414*	0.0603	0.0319	0.1393*
Unemployment	-0.2479*	-0.2048*	-0.3938*	-0.4503*	-0.3589*

Table 8c. Pearson correlation with final estimate GDP gap

	1995-99	2000-04	2005-09	2010-14	2015-18
Inflation	0.0782	-0.1392*	0.063	0.1128*	-0.0102
CA	-0.2021*	0.1077*	0.0179	0.1836*	-0.0529
Unemployment	-0.0243	0.2053*	-0.0282	-0.3617*	-0.1787*

Table 8d. Spearman correlation with final estimate GDP gap

	1995-99	2000-04	2005-09	2010-14	2015-18
Inflation	0.1690*	-0.1023*	0.2592*	0.2953*	0.1179*
CA	-0.2365*	-0.0382	-0.2566*	0.0284	0.0248
Unemployment	-0.1018	-0.0837	-0.1584*	-0.3466*	-0.2389*

Turning to income level, in Table 9 below we compare the basic results across country types. We use the same *World Economic Outlook* classification as in the previous section for advanced economies, emerging markets, and low-income countries to gauge differences across income level³². Again, we see some evidence across country groups of a relationship

³² We restrict ourselves to real time as opposed to final output gap estimates for this exercise. We continue to use yearly changes to focus on the movement, as opposed to the level of the variables. Note that for low income countries, there is limited data as we only have 49 total observations that all fall in the 2010-2018 period.

that goes with the economic intuition. Inflation is weakly correlated with output gaps across country groups, while unemployment is sizeable and negatively correlated with gaps, except in LICs. This result holds when looking at leads and lags. This suggests that output gaps at time t provide some, albeit imperfect, information about concurrent demand pressures or developing ones.

Table 9. Real time GDP gaps: correlations with other measures of slack, 2009-18

Variables	Pearson		Spearman	
	AE	EM& LIC	AE	EM&LIC
Inflation	0.0908*	0.0002	0.1057*	0.0241
CA balance	0.2462*	0.1384*	0.3056*	-0.0149
Unemployment rate	-0.5783*	-0.2042*	-0.5658*	-0.3142*

The above exercises are functionally similar in scope and sophistication to the Phillips curve. Therefore, for robustness, we also estimate a few simple variations of Phillips curve that take some variant of the following form³³:

$$\Delta\pi_t = \alpha + \beta_i \Delta GAP_t + \varepsilon_t$$

Before turning to the results, it should be noted that the large sample prevents us from including an expectation component and the exercise thus reverts to one of simple co-movement. While an expectation-augmented Philips curve would help shed more light on the relationship between nominal variables and economic activity, our specification can add weight to the results in the previous exercises. In Table 10 below, the results of these simple estimates, using various combinations of lags and leads, as well as for different periods and subsamples, confirm that the relationship between the two has the right sign but the effects are partial. A select few specifications show a significant β -coefficient, but its size is small, and the overall fit is miniscule.

To reiterate, the relationship between inflation and output gaps is imperfect and empirical estimates would ex ante be expected to be volatile. This is not least due to the fact that many other factors affect inflation, including exchange rate movements and cost push shocks. One would therefore expect a positive, albeit imperfect, unconditional relationship between the two, as we do indeed see in these simple empirical results. Also, as expected, the fit of the model improves if we use revised output gap data, but the R-squared values are still small, even in the case of panel data regression³⁴. This is in line with Banbura and Bobeica (2020)

³³ It would be preferable to include an expectation component in the above but given our sample of over 100 countries over a 40-year period, this is not feasible due to data limitations.

³⁴ To reiterate, this very partial relationship can be due to several factors. In fact, a correctly estimated output gap, in some sense, could still show a limited but positive relationship with inflation if the true underlying relationship between slack and price pressures had weakened or is non-linear, as has been extensively discussed in the recent literature. See e.g. Del Negro et al (2020).

who find that the Phillips curve for the euro area does improve inflation forecasts for the euro area but that the relationship is imperfect and specification choices are important.

Table 10. Regressions of the change in inflation of measures of changes in output gap

Variables	(1) Pooled OLS	(2) Panel FE	(3) Panel RE	(4) Pooled OLS	(5) Panel FE	(6) Panel RE
Delta Real-time Output Gap	0.229 (0.197)	0.200*** (0.0685)	0.200*** (0.0685)			
Lagged Delta Real-time Output Gap	0.0205 (0.167)	0.122** (0.0586)	0.122** (0.0586)			
Delta Real-time Output Gap (2 lags)	-0.193 (0.142)	0.0767 (0.0503)	0.0767 (0.0503)			
Delta Final Output Gap				0.301*** (0.0295)	0.315*** (0.0308)	0.315*** (0.0308)
Delta Lagged Final Output Gap				0.0850*** (0.0270)	0.0975*** (0.0283)	0.0975*** (0.0283)
Delta Final Output Gap (2 lags)				-0.0243 (0.0253)	-0.0196 (0.0276)	-0.0196 (0.0276)
Constant	0.413 (0.297)	0.383*** (0.101)	0.383*** (0.101)	-0.171** (0.0742)	-0.166** (0.0746)	-0.166** (0.0746)
Observations	803	803	803	807	807	807
R-squared	0.005	0.016	0.016	0.134	0.148	0.148

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 11 shows the results of identical regressions for core inflation.³⁵ The results are broadly similar to the one in Table 10, indicating that the relationship between the output gap and inflation is a robust one and supports a Philips curve-type link between inflation and the output gap.

³⁵ The group of countries for which comparable core inflation data are available is Australia, Belgium, Canada, Cyprus, Denmark, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, Malta, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom.

Table 11. Regressions of the change in core inflation of measures of changes in output gap

Variables	(1) Pooled OLS	(2) Panel FE	(3) Panel RE	(4) Pooled OLS	(5) Panel FE	(6) Panel RE
Delta Real-time Output Gap	-0.171* (0.103)	- 0.148*** (0.0415)	- 0.148*** (0.0415)			
Lagged Delta Real-time Output Gap	-0.0103 (0.0954)	-0.0274 (0.0384)	-0.0274 (0.0384)			
Delta Real-time Output Gap (2 lags)	0.0117 (0.0704)	0.000922 (0.0287)	0.000922 (0.0287)			
Delta Final Output Gap				-0.119 (0.0922)	- 0.124*** (0.0373)	- 0.124*** (0.0373)
Delta Lagged Final Output Gap				0.0434 (0.0918)	0.0357 (0.0371)	0.0357 (0.0371)
Delta Final Output Gap (2 lags)				0.0779 (0.0905)	0.0438 (0.0367)	0.0438 (0.0367)
Constant	2.014*** (0.158)	2.012*** (0.0634)	2.012*** (0.0634)	1.987*** (0.151)	1.988*** (0.0609)	1.988*** (0.0609)
Observations	427	427	427	446	446	446
R-squared	0.007	0.031	0.031	0.007	0.034	0.034

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

IV. OUTPUT GAP COVERAGE IN IMF BILATERAL SURVEILLANCE

Within its surveillance function, the IMF produces Staff Reports for its membership, typically on an annual basis. These reports provide a comprehensive review of member countries, including an assessment of each country's economy, its policies, and recommendations. In this section, we shed further light on the use of the output gap in this surveillance practice. We use text analysis methods to document how output gaps are discussed and covered in Staff Reports. Our sample includes 195 countries covering the period 2000-2019, providing a total of 2536 Staff Reports.³⁶ As a robustness check, we also manually analyze 160 Staff Reports for 12 countries in an attempt to look more closely at the context and framing of output gap coverage for a select group of countries.

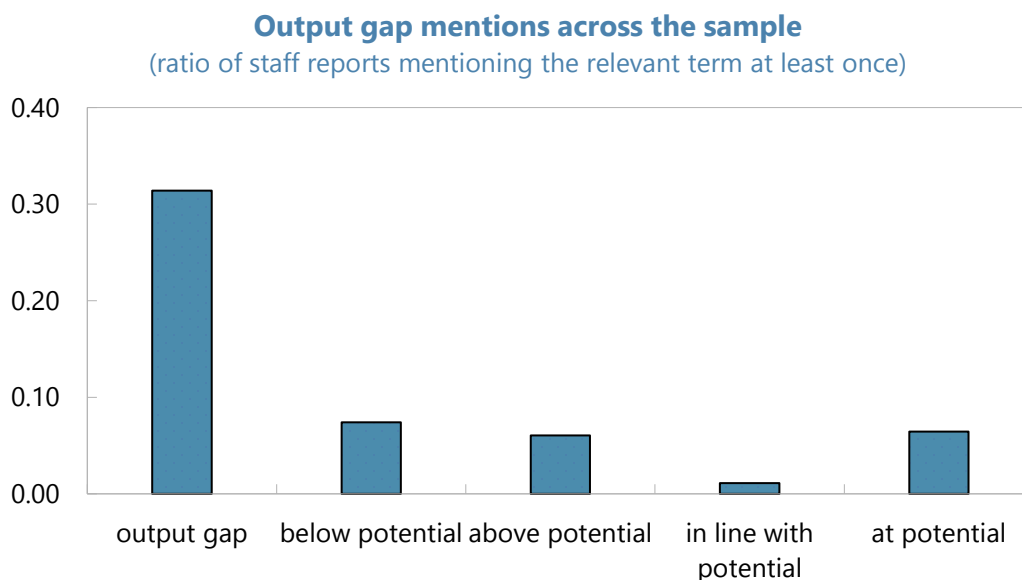
There is a growing literature using natural language processing (NLP) techniques to assess the discussion of a certain topic in IMF outputs. These techniques can help analyze the extent and context of coverage in reports as well as their development over time. Duval and Loungani (2019) employ text mining techniques to evaluate IMF recommendations in the areas of employment protection, unemployment insurance, and minimum wages. Fayad et al.

³⁶ A number of staff reports are not available for text analysis. This is the case when reports are confidential or not properly documented in the IMF's institutional repository.

(2020) analyze traction in Article IVs by assigning macroeconomic topics and sentiment to authorities' views paragraphs. Cherif, Engher, and Hasanov (forthcoming) use text analytics to identify growth narratives in IMF Staff Reports over a sample spanning 40 years.

Figure 1 below shows an initial summary of output gap coverage for our text analysis sample group. The terms we use for the search are “output gap”, as well as text describing the

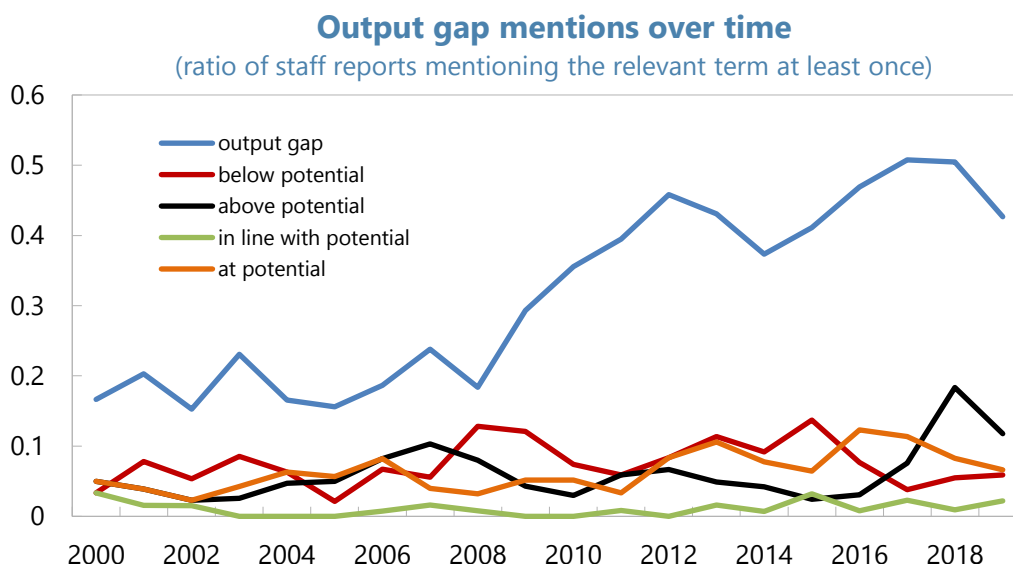
Figure 1.



relationship between actual and potential growth. These are “above potential”, “in line with potential”, “at potential”, and “below potential”. These terms should strictly relate to the output gap as its definition is the deviation of actual growth from its potential. We exclude the more general term of “potential growth” on its own as it is used regularly in discussion of structural reforms to raise medium term growth and thus not directly related to the assessment of the economy’s current cyclical position. Similarly, other terms such as “slack” were deemed too broad and were frequently raised in context unrelated to the output gap itself.

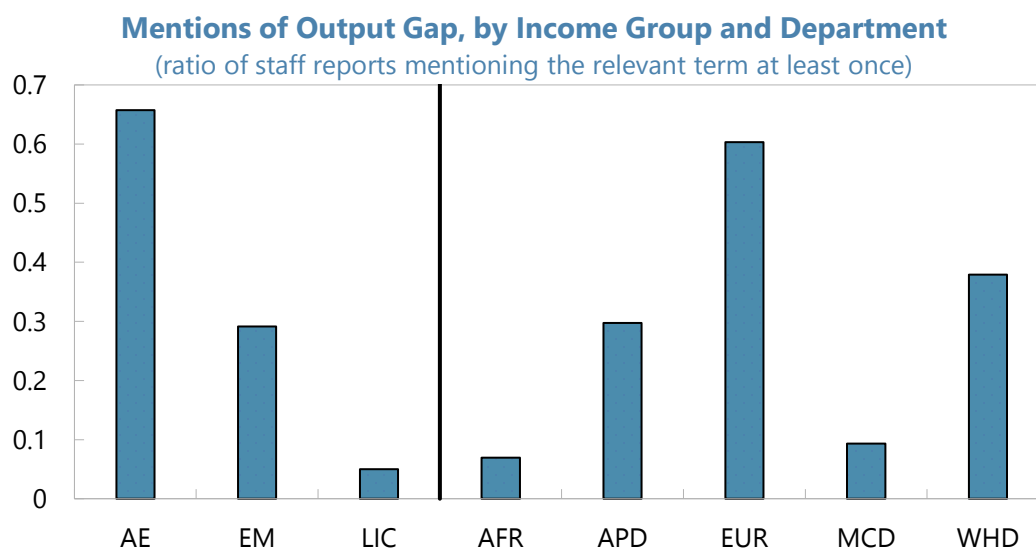
Across our sample, 31% of Staff Reports discuss the output gap at least once. As Figure 2 shows, the number of reports mentioning the output gap has increased over time, with a notable spike following the Global Financial Crisis. Staff Reports covering the terms “below potential” and “at potential” also show an upward trend over time, albeit with a sizable standard deviation. “Below potential” experienced significant coverage around the GFC, but levels off after that, while “above potential spikes up dramatically in the last year of our sample. As mentioned in previous sections, this could be due to the general assessment of closing output gaps in many countries and diminishing slack prior to the outbreak of Covid-19. The last term we searched for, “in line with potential”, remains relatively flat across time. It is noteworthy that the level of coverage for the term “output gap” is much higher than that of the other terms.

Figure 2.



Analyzing the coverage across country groups yields intuitive results: 66% of Staff Reports covering Advanced Economies mentioned the output gap, versus 29% for Emerging Markets, and only 5% of Low-income Countries³⁷. The other terms follow similar patterns. This suggests that cyclical considerations are more prominent in more advanced economies whereas structural issues may dominate more for developing countries. Despite significant disparities between number of Staff Reports covering output gaps across income groups, the trends over time are similar: coverage is stagnant or even decreasing in the first half of our sample, and quickly expands following the GFC. Looking at regional distinctions, we find

Figure 3.



³⁷ These figures apply to use of the term “output gap” but not the other terms discussed in the previous paragraph.

that countries in Europe (60%), the Western Hemisphere (38%), and Asia/Pacific (30%) are most likely to discuss the output gap in their Article IVs.

Following these basic results, we juxtapose the above calculations with numerical estimates of the output gap but find no conclusive relationship. That is to say, output gaps do not seem to be discussed more in countries where the absolute gap is estimated to be relatively larger. The median real-time estimate output gap is not statistically different for countries and in years in which they discuss the output gap in their Staff Report.³⁸ In addition, we compare textual references to the output gap to annual changes in the output gap estimates as well as size of revisions, but arrive at the same conclusion.

The relationship between the discussion of output gaps in a Staff Report and a country's output gap estimate does not have to be contemporaneous, however. We therefore analyze the median output gaps, annual changes, and revisions for countries and in years for which the Article IV mentions the output gap either in the year prior, or in the year after. The logic here is that a country team could either predict the appearance of a relatively large output gap beforehand, or evaluate it post hoc. However, we found no statistically significant difference in results between countries output gaps in $t-1$ or $t+1$ and those that did not.

Specific country characteristics, beyond the size of their output gap, could also spur a discussion of the output gap in a country's Staff Report. To assess this, we run a simple probit model with the dependent variable being a dummy signifying whether country i mentions the term 'output gap' in year t , and a number of macroeconomic variables, sectoral risk measures, and IMF policy advice. We also include a lagged dependent variable, accounting for a potential serial correlation in output gap mentions. The model controls for time effects where necessary.³⁹

The specification containing country characteristics of a macroeconomic nature can be found in column (1) of Table 12. We include log of GDP per capita, growth, inflation, and unemployment. GDP has a large, positive coefficient and is significant, which is broadly in line with the aforementioned finding that advanced economies have a higher tendency to bring up output gaps in their Staff Reports. Furthermore, the coefficient on inflation is negative and significant, suggesting that output gaps are discussed more often during periods of disinflation or lowflation. Conversely, the results suggest a positive relationship between growth and the output gap mentions. The balance of payments has a significant, negative impact on output gap mentions, while unemployment bears no clear effect.

The specification in column (2) contains a number of sectoral risk measures. As the table demonstrates, the coefficients on External risk and Fiscal risk are negative and significant, suggesting that countries that face less external risk and/or fiscal vulnerabilities are more likely to discuss output gaps in their Article IV. This may also relate to the finding, discussed

³⁸ We normalize the output gap estimates to account for any country-specific characteristics.

³⁹ We also run robustness checks controlling for country-fixed effects. This does not qualitatively change the results but removes significance for the variables Current Account Balance and the risk measures.

above, that output gaps are more often discussed in the context of advanced economies, which inherently tend to exhibit lower external risks than emerging market economies and low-income countries, while also exhibiting a relatively lower probability of sovereign default.

To look more closely at IMF surveillance and output gap estimates, in the following we regress mentions of the output gap on Fund policy advice, i.e. the degree to which the IMF recommended the country in question to tighten (loosen) their fiscal or monetary stance⁴⁰. None of the variables bear any significance, suggesting no apparent relationship between Fund advice and a country team's inclination to discuss the output gap, i.e. cyclical considerations do not impact IMF policy advice substantially. However, data for policy advice was only available for two years, so the sample is much smaller than for the other specifications. It should also be noted that the recommendations are in relation to actual policy. Therefore, they suggest to what extent policy should be eased or tightened, in comparison to the actual policy stance of the authorities, as opposed to whether the policy should be tight or loose in the abstract.

⁴⁰ This exercise uses an internal database which logs the policy recommendation of each country team, including on monetary and fiscal policy.

Table 12. Output gap mentions and country characteristics

Variables	(1) Mentions of Output Gap in SR	(2) Mentions of Output Gap in SR	(3) Mentions of Output Gap in SR	(4) Mentions of Output Gap in SR
Lagged Mentions of Output Gap in SR	1.120*** (11.87)	1.177*** (13.13)	1.883*** (7.64)	1.932*** (5.28)
Real-time Output Gap - Absolute				
GDP in Capita in USD in logs	0.244*** (4.84)			
Real GDP Growth	0.0188 (1.14)			
Inflation	-0.0234* (-2.35)			
Unemployment	0.00170 (0.19)			
Current Account Balance	-0.0142** (-2.68)			
External Risk		-0.0187*** (-6.33)		
Fiscal Risk		-0.0099*** (-3.34)		
Financial Risk		0.00977** (2.86)		
Monetary Recommendation			-0.00512 (-0.06)	
Fiscal Recommendation			-0.135 (-1.25)	
Change in Monetary Recommendation				0.139 (0.87)
Change in Fiscal Recommendation				-0.0666 (-0.39)
Constant	-2.721*** (-5.15)	-0.275 (-1.26)	-0.388 (-0.88)	-0.965*** (-3.35)
Observations	1051	1370	153	73
R-squared				
Time FE	Yes	Yes	No	No
Standard errors in parentheses				

Another slice that allows us to gauge the relationship between the state of the business cycle and policy recommendations is provided in the figure below. Figure 4 shows how recommendations vary depending on the level of the output gap. We look at both the level of the output gap compared to the recommendation, as well as the change in the output gap and the change of the recommendation. We see a slight positive association between the level of

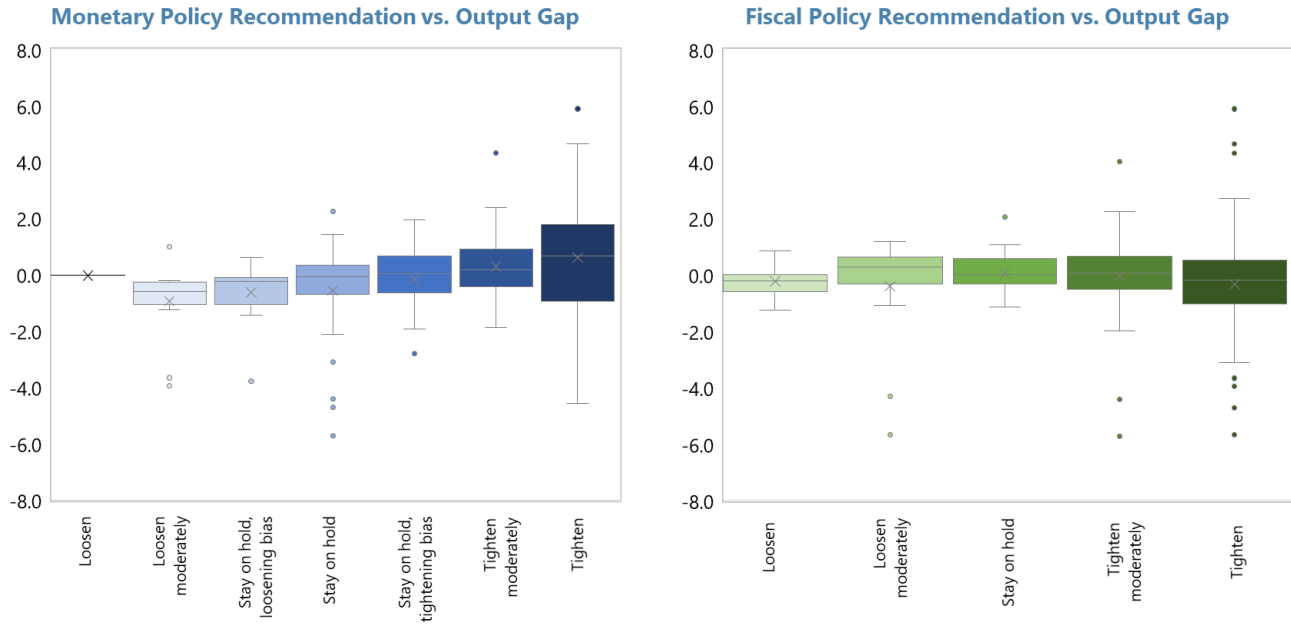
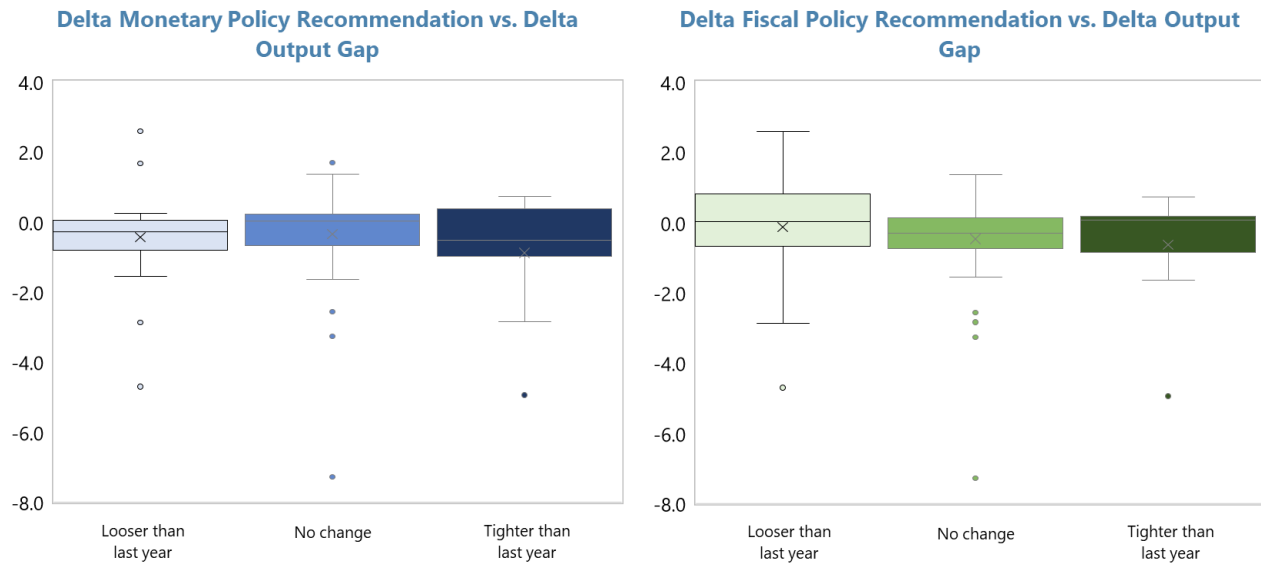
the output gap and the recommended relative tightening of monetary policy but a very limited trend for the other variations.

The lack of a more robust relationship between the output gap and recommended stance could be driven by several factors. First, as mentioned above, the recommendations are in comparison to the actual policy stance of the authorities, which should already to some extent incorporate the status of the output gap. Therefore, if a country has, for example, already eased policy in response to a negative output gap, the recommended policy stance in the Staff Report would take this into account and thus not necessarily have an easing bias.

Another important factor are that other considerations might prevent a recommended countercyclical policy action. One of the most obvious examples of this would apply to fiscal advice in the presence of fiscal risks whereby a Staff Report could not be able to recommend easier fiscal policy for a country with a high level of public debt. To test this additional constraint, we look at the simple relationship between the fiscal recommendation and the level of the output gap, conditional on our fiscal risk variable. While such a regression excludes other country-specific characteristics and is only indicative, we do find that both the output gap and fiscal risk are significant variables with coefficients in line with expectations, although the level of variation explained by the variables is limited.

An equivalent look at whether external risks may prevent a recommended monetary loosening yields somewhat different results. The external risk variable is highly significant with the right sign and slightly higher coefficient than that of fiscal risk for fiscal recommendations. That is to say, reports for countries that exhibit a high level of external risk are less likely to recommend an easing of monetary policy. This could be due to concerns over exchange rate considerations and the possibility that lower rates could trigger capital outflows. However, the size of the output gap itself becomes non-significant, suggesting that factors other than the state of the business cycles, such as external risks, are more important for the monetary policy recommendation.

The overall pattern from the above is thus one whereby there is a positive but weak relationship between policy recommendations and the level of the output gap. The lack of a stronger relationship can partially be explained by obvious constraints on policy, such as capacity constraints on the government's balance sheet and external risks. Future research could delve more deeply into country characteristics and examine whether policy recommendations are indeed robustly determined by output gaps once most plausible underlying factors and constraints are taken into account.

Figure 4a. Policy Recommendations vs. Output Gap**Figure 4b. Change in Policy Recommendations vs. Change in Output Gap**

In addition to the above, text analysis also allows us to gauge the context in which output gaps are mentioned in Staff Reports. Following the methodology introduced by Fayad et al. (2020), we assign

macroeconomic topics to each paragraph in a Staff Report, using a vector of terms associated with those topics. The topics are: External Sector, Financial Sector, Fiscal Sector, Monetary Sector, and Real/Structural Sector. In line with Fayad et al, we use the IMF’s Enterprise Business Vocabulary (EBV), which contains the aforementioned sectors and a number of terms related to those sectors. Though the structure of the EBV allows us to easily assign macroeconomic terms to each sector, the language used is often abstract and the exact words do not generally appear in Staff Reports.

Due to this, we enter the the terms associated to each sector in a Word2Vec model (Mikolov et al., 2013) trained on IMF Staff Reports. Word2vec generates a vector space, typically of several hundred dimensions, with each unique word in our corpus of Article IVs being assigned a corresponding vector in the space. Word vectors are positioned in the vector space such that words that share common contexts in the corpus are located close to one another in the space. We extract the terms most similar to our five sectors. Lastly, we match these terms with the contents of individual paragraphs and determine to which sector the paragraph corresponds to most. Fayad et al. find that this method achieves an 88% accuracy rate, meaning that in 88% of the cases, this approach assigns the same topic a human would. However, it should be stressed that the potential for errors remains and we therefore interpret the results with caution.

Looking at the initial results for our sample, 36% of paragraphs discuss the Real sector, followed by Fiscal (29%), Financial (20%), External (18%), and Monetary (6%).⁴¹ As can be seen in Table 13, we find that the majority of mentions of output gap occurs in the context of the Real Sector, namely 64%. A necessary disclaimer here is that the dictionary described above does include the term “output gap” as being similar to the Real Sector, so there is some bias here. Nonetheless, discussing the output gap along within the real sector makes intuitive sense as this is the part of the report that discusses macroeconomic developments and trends. Furthermore, we see that references to the output gap are more likely to occur in discussions of the Monetary Sector (20%) than the Fiscal Sector (16%). This finding is particularly salient as only 6% of total paragraphs in our sample are assigned to the monetary sector, while 29% of all text in Article IVs discusses fiscal matters. This is intuitive, given monetary policy’s preeminence as the business cycle stabilization tool of choice in recent decades. In comparison, a mere 6% of mentions take place in the context of the External Sector, and only 2% within discussions of the financial sector.

The other terms describing the relationship between actual and potential growth display similar patterns. Noteworthy is that the phrases “above potential” and “in line with potential” are almost exclusively featured in paragraphs covering real sector issues. Furthermore, an added distinction is that though sections discussing the monetary sector frequently mention the output gap, occurrences of the other terms are much rarer.

⁴¹ Note that the sum of these shares surpasses a 100%, as we allow double assigning whereby a paragraph can discuss two topics equally.

Table 13. Output gap mentions by sector

	Over total # of paragraphs	Over total # of paragraphs mentioning:	"Output gap"	"Below potential"	"Above potential"	"In line with potential"	"At potential"
Over total # of paragraphs by sector:	100%		1.2%	0.2%	0.1%	0.0%	0.1%
External	18%		6%	11%	8%	3%	7%
Financial	20%		2%	6%	0%	3%	13%
Fiscal	29%		16%	19%	22%	9%	14%
Monetary	6%		20%	13%	4%	6%	3%
Real	36%		64%	61%	72%	85%	68%

Despite the absence of a precise relationship between the output gap and the other indicators of slack in the previous section, we do find that Staff Reports will at times mention these measures in the same breadth, as reported in Table 14 below. To start, we find that ‘inflation’ is mentioned at least once in 12.4% of all paragraphs in our sample, followed by current account balance ⁴² (6.5%), and ‘unemployment’ (3.5%). Looking at co-occurrences of these phrases and the output gap, we find somewhat unsurprisingly that 55% of paragraphs across our sample that discuss the output gap, also mention ‘inflation’. Indeed, 95% of all monetary paragraphs covering the output gap also touch on inflation. ‘Unemployment’ surfaces in 15% of paragraphs that consider the output gap, mostly in the context of the real sector, followed by current account balance at 13%. For comparison, we also look at capacity utilization which turns out to be less likely to overlap with references to the output gap, at 4%⁴³.

Table 14. Output gap co-occurrences with measures of slack

	"Unemployment"	"Inflation"	"Current Account Balance"	"Capacity Utilization"
Over total # of paragraphs	4%	12%	7%	0.2%
Over total # of paragraphs mentioning "output gap" by sector:	15%	55%	13%	4%
External	1%	3%	3%	0%
Financial	0%	1%	0%	0%
Fiscal	1%	2%	0%	0%
Monetary	0%	18%	1%	0%
Real	14%	35%	9%	4%

Overall, these results confirm that the output gap is a widespread concept in bilateral IMF surveillance. Its use has increased over time and its relevance is greater for more advanced countries, and on a sectoral basis when discussing the real economy and monetary policy.

⁴² The search terms for “current account balance” also include “current account”, “current account deficit”, and “current account surplus”.

⁴³ We ran a separate exercise for the “Authorities’ Views” section of Staff Reports which present the authorities’ represent the part of the reports where authorities lay out their view of the analysis and to the extent to which they agree with Fund staff. These views are laid out in single paragraphs for each section and are thus much shorter than staff’s analysis. We therefore see that, as expected, the number itself is much less frequent than for the main text. The average paragraph discusses output gaps between 1-5% of the time, with some increase in recent years. Of note, the monetary policy section enjoys the highest frequency of output gap coverage.

The simple empirical exercises showed the presence of serial correlation in output gap coverage, but that, according to our methods, there is no apparent relationship between the size of the gap itself and the extent of its coverage.

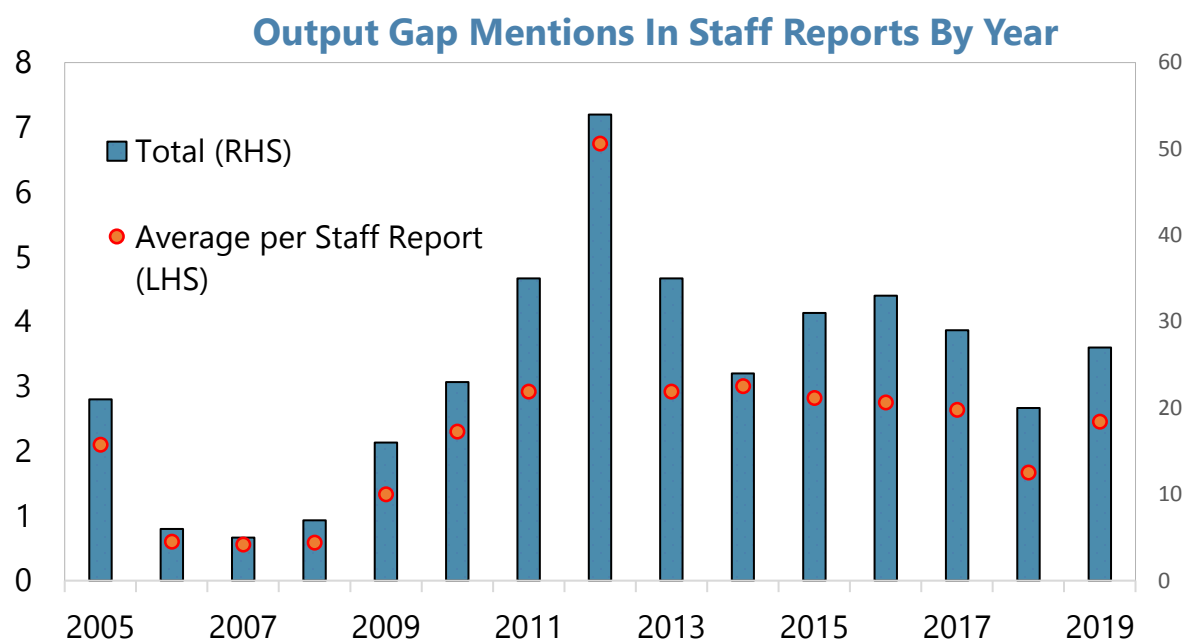
To complement the formal text analysis exercise and provide a cross-check on the results, we also perform some manual text analysis by going through a select sample of Staff Reports for 12 countries. Such a cross-check can provide more nuance to the more systematic analysis above, and potentially add more depth to the context within which output gaps are discussed. The countries chosen are the G-7 members, Greece, Ireland, Portugal, Spain, and Korea, resulting in a total of 160 reports. We manually search for the same terms as in the text analysis above, noting the frequency and location by section.

Figure 5 below shows the raw number and proportion of times that output gaps are mentioned over time⁴⁴. This is broadly consistent with the text analysis above, showing that the output gap is a prominent topic of discussion in staff reports and, in fact, increasingly so during and following the global recession. Looking at individual years, there is a clear peak in 2012, coinciding with the euro area crisis although attributing it to the crisis would require more formal text analysis⁴⁵.

⁴⁴ This excludes boxes and annexes which can skew the aggregate numbers significantly upward when they focus on the topic in question.

⁴⁵ These results hold when looking at unique mentions per Staff Report. For example, one report could include multiple references while others for the same year might include no references at all, complicating the interpretations of average per year. When we look only at whether a given report includes a mention or not, however, the same trend holds. In turn, the number of reports with no mention shows no clear trend.

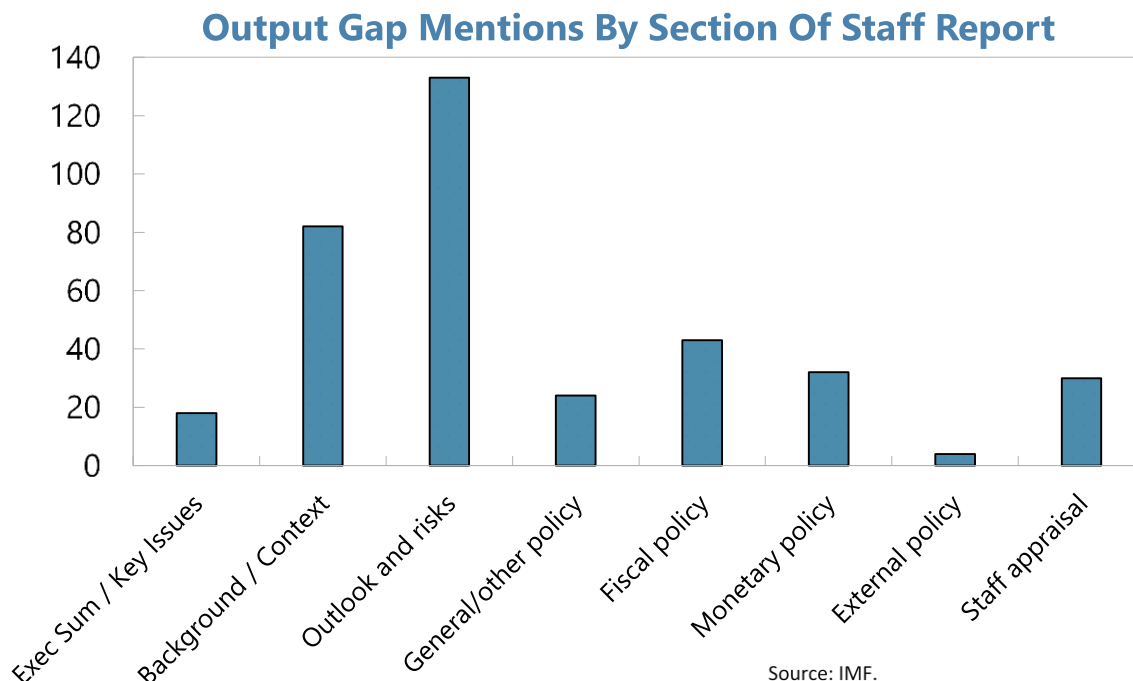
Figure 5.



Source: IMF.

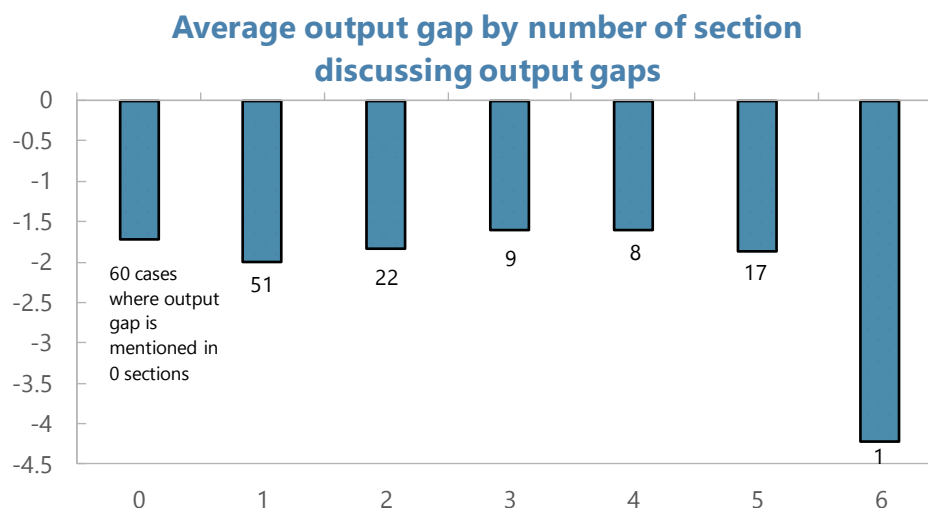
In terms of context of the coverage, Figure 6 plots the section within which output gaps are discussed. It should be noted that the title of sections is not strictly homogenous, although it has become more so over time. Furthermore, the categorization is different to that of the formal text analysis above. For the manual analysis we attribute sections to their actual subject as opposed to the attribution of paragraphs in the systematic text analysis. The staff reports in our sample show a large similarity in terms of titles of sections within Staff Reports which allows us to compare coverage both across time and between countries. The chart shows that output gaps are most often mentioned in the forward-looking general section of the reports, usually titled “Outlook and risks”. The bulk of the material in this section would be attributed to the real sector in the formal text analysis, which also saw significant coverage of output gaps. There is also considerable coverage in the backward-looking context section, titled “Background / Context”, while on a policy basis, output gaps are mentioned roughly in equal measure when it comes to fiscal policy and monetary policy. This doesn’t fully align with our formal text analysis, where monetary policy has a higher coverage than fiscal policy. However, the difference in coverage between the two policy sections in the manual analysis is quite limited.

Figure 6.



Finally, in Figure 7 below, we look at our manually collected sample of output gaps and compare it to the actual estimated output gap for the given country-year. We look at what the average output gap is for each Staff Report that mentions output gaps in a certain number of sections. For example, the first column shows that the average output gap is -1.7% in Staff Reports that do not mention output gaps (60 such cases). Similarly, the average gap is -2% for the 51 reports that mention gaps in one section. As can be seen in the chart, there is no apparent correlation between the estimated gap itself and the intensity of the coverage of output gaps. For robustness, we also looked at the frequency of coverage, as opposed to the number of sections involved, and similarly find no trend.

Figure 7.



More research is needed in this area going forward. In particular, the formal text analysis above could be extended to encompass the relationship between output gap and policy advice or sentiment demonstrated in the discussion of output gaps. Such an exercise would also help compare these instances to the output gap to gauge the extent to which greater demand pressure coincide with calls for tighter policy.

V. CONCLUDING THOUGHTS

Countercyclical macroeconomic policy relies crucially upon a reliable real-time assessment of an economy's cyclical position. In practice, an important part of this assessment has been an estimate of the economy's growth momentum compared to its hypothetical potential. The gap between the two is at all times uncertain but their imperfectly estimated levels help guide policymakers and economists towards the assessment of the cyclical state of the economy. In this paper, we have provided an overview of how the output gap is used in IMF surveillance along three main channels: historical estimates of output gaps, the relationship between output gaps and other variables of interest, and IMF coverage of output gaps in its annual Staff Reports.

The historical evidence on estimated output gaps for the world economy shows that they have a negative skew, meaning that their average over time is not zero. This is in line with previous findings in the literature for particular regions, most notably Europe. Assuming that the length of time we consider and the number of countries assure that recessions are not overrepresented in the data, this suggests an assessment that the world economy has performed under its potential to a certain degree over time⁴⁶. Several theoretical reasons for

⁴⁶ It is, conversely, possible that, despite the large set of countries and relatively long sample period, the data used in preceding sections may not be large enough to rule out sample bias. While we do slice the data into pre- and post-GFC periods, the strength of the business cycle pre-GFC may not have been remarkable while the
(continued...)

such a state of affairs have been proposed, including the effect of nominal wage rigidities (Aiyar and Voigts (2019) and a plucking framework for business cycles (Dupraz et al (forthcoming)). Such hypotheses have gained in popularity in recent years, not least in light of limited signs of overheating for long periods and the fact that a symmetric view of business cycles is hard to square with recent experience.

If the negative skew conversely represents a true measurement error, the argument can be made that the degree of slack in the world economy has been less than estimated. This could lead to policy advice being too loose and recommending too accommodative policies. However, as noted above, the supporting evidence does not wholly support this view as other measures of slack, notably inflation, have not suggested that production has generally exceeded its capacity constraints in many countries.

In addition to the negative skew, the historical data shows frequent revisions to estimates of potential output and consequently to the assessment of the output gap. These changes are only very partially explained by revisions to the underlying realized output data, and therefore likely reflect changing views on the economy's underlying fundamentals, reinforcing the inherent difficulty of real-time assessment. These main findings for the historical estimates broadly hold across different countries, income levels, and time periods.

The second core section of our paper documented the relationship between output gap estimates, on the one hand, and measurable indicators of slack, on the other. We found limited but nonetheless significant correlation between the two in general. Furthermore, the signs go in the direction that one would expect from economic theory. However, the co-movement between output gaps and other measures of slack is strengthened by the use of final estimates of output gaps, which strengthens the argument for using caution in overreliance on real-time estimates of output gap in policymaking.

Our review of Staff Reports showed that output gaps figure prominently in IMF country surveillance. This is the case both in our formal and manual text analysis. The incidence of coverage seems to have increased over time, most notably following the Global Financial Crisis. The incidence is also greater for more advanced economies, and figures most heavily in discussions of developments in the real economy, as opposed to policy sections. Finally, it does not seem to be the case that output gaps are discussed more prominently in cases where the gap itself is larger.

There are no obvious silver bullets that address the shortcomings and challenges of output gap estimation, and in particular their interpretation in real time, as described above. One possible improvement, however, would be to explicitly acknowledge the profound uncertainty of the estimates, especially in real time. The relevance of accepting possibly higher margins of error around point estimates of output gaps will also be more pronounced in the post-Covid 19 environment given the effects of the crisis on both demand and supply of output. In practice, this would involve an emphasis on risks in discussion of slack

post-GFC recovery could have been usually slow and weak. On the latter, Duval et al (2019) provide evidence of hysteresis effects from financial frictions following the GFC in advanced economies.

estimates and placing policy discussions in the context of contingent macroeconomic developments to a greater extent. Furthermore, more frequent use of confidence intervals, as opposed to point estimates which are the overwhelming norm, would serve to highlight the lack of concrete knowledge and support more contingent assessment and policy advice. This is in line with Romer (2020) who documents a similar overreliance on point estimates in the broader empirical academic literature.

REFERENCES

- Aiyar, S. and Voigts, S., 2019, “The Negative Mean Output Gap”, IMF WP 19/183.
- Banbura, M., and Bobeica, E., 2020, “Does the Phillips curve help to forecast euro area inflation?”, ECB Working Paper Series, No 2471 / September 2020.
- Bekaert, G., Engstrom, E. and Ermolov, A., 2020, Aggregate Demand and Aggregate Supply Effects of COVID-19: A Real-time Analysis, Finance and Economics Discussion Series 2020-049, Board of Governors of the Federal Reserve System.
- Benes, J., Clinton, K., Garcia-Saltos, R., Johnson, M., Laxton, D., Manchev, P. and Matheson, T., 2010, “Estimating Potential Output with a Multivariate Filter”, IMF Working Paper 10/285.
- Benigno, G. and Fornaro, L., 2017, “Stagnation Traps”, *The Review of Economic Studies*, Vol. 85(3), pp. 1425-1470.
- Berger, H., Dowling, T., Lanau, S., Lian, W., Mrkaic, M., Rabanal, P. and Taheri Sanjani, M., 2015, “Steady as She Goes— Estimating Potential Output During Financial “Booms and Busts”, IMF Working Paper 15/233.
- Blagrove, P., Garcia-Saltos, R., Laxton, D. and Zhang, F., 2015, “A Simple Multivariate Filter for Estimating Potential Output”, IMF Working Paper 15/79.
- Borio, C., Disyatat, P. and Juselius, M., 2017, “Rethinking Potential Output: Embedding Information about the Financial Cycle”, *Oxford Economic Papers*, Vol. 69(3), pp. 655–677.
- Brooks, R. and Basile, G., 2019, “Campaign Against Nonsense Output Gaps (CANOO)”, IIF Global Macro Views.
- Buti, M., Carnot, N., Hristov, A., McMorro, K., Roeger, W. and Vandermeulen, V., 2019, “Potential Output and EU Fiscal Surveillance”, VoxEU Blog Post.
- Cerra, V. and Saxena, S., 2017, “Booms, Crises, and Recoveries: A New Paradigm of the Business Cycle and Its Policy Implications”, IMF Working Paper 17/250.
- Chen, J. and Gornicka, L., 2020, “Measuring Output Gap: Is It Worth Your Time?”, IMF Working Paper 20/24.
- Cherif, R., Engler, M. and Hasanov, F., forthcoming, “Crouching Beliefs, Hidden Biases: The Rise and Fall of Growth Narratives, IMF Working Paper.
- Claus, I., 2000, "Is the Output Gap a Useful Indicator of Inflation?", Reserve Bank of New Zealand Discussion Paper Series DP2000/05.

Coibion, O., Gorodnichenko, Y. and Ulate, M., 2018, “Real-Time Estimates of Potential GDP: Should the Fed Really Be Hitting the Brakes?”, Center on Budget and Policy Priorities.

Del Negro, M., Lenza, M., Primiceri, G. and Tambalotti, A., 2020, “What's Up with the Phillips Curve?”, NBER Working Paper No. 27003.

Dupraz, S., Nakamura, E. and Steinsson, J., forthcoming, “A Plucking Model of Business Cycles”, Revise and resubmit at Journal of Political Economy.

Duval, R., Hong, G. and Timmer, Y., 2019, “Financial Frictions and the Great Productivity Slowdown”, Review of Financial Studies, Vol. 33 (2), pp. 475-503.

Duval, R. and Loungani, P., 2019, “Designing Labor Market Institutions in Emerging Market and Developing Economies: Evidence and Policy Options”, IMF Staff Discussion Note 19/04.

Eggertsson, G., Mehrotra, N. and Robbins, J., 2019, “A Model of Secular Stagnation: Theory and Quantitative Evaluation”, American Economic Journal: Macroeconomics, Vol. 11(1), pp. 1–48.

Fayad, G., Huang, C., Shibuya, Y. and Zhao, P., 2019, “How Do Member Countries Receive IMF Policy Advice: Results from a State-of-the-art Sentiment Index”, IMF Working Paper 20/7.

Gordon, R., 1997, “The Time-Varying NAIRU and its Implications for Economic Policy”, Journal of Economic Perspectives, Vol. 11 (1), pp. 11-32.

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”, Review of Economics and Statistics, Vol. 100(5), pp. 831–843.

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

Hooper, P., Mishkin, F. and Sufi, A., 2019, “Prospects for Inflation in a High Pressure Economy: Is the Phillips Curve Dead or is It Just Hibernating?”, NBER Working Paper No. 25792.

Hristov, A., Raciborski, R., and Vandermeulen, V., 2017, “Assessment of the Plausibility of the Output Gap Estimates”, European Commission Economic Brief 023.

Independent Evaluation Office of the IMF, 2014, “IMF Forecasts: Process, Quality, and Country Perspectives”, Evaluation Report, ISBN 978-1-47559-951-0.

International Monetary Fund, 2019, “World Economic Outlook, October 2019 Global Manufacturing Downturn, Rising Trade Barriers”, ISBN 978-1-51351-616-5.

Jarocinski, M. and Lenza, M., 2018, “An Inflation-Predicting Measure of the Output Gap in the Euro Area”, *Journal of Money, Credit and Banking*, Vol. 50(6), pp. 1189-1224.

Kangur, A., Kirabaeva, K., Natal, J. and Voigts, S., 2019, “How Informative Are Real Time Output Gap Estimates in Europe?”, IMF Working Paper 19/200.

Kuh, E., 1966, “Measurement of Potential Output”, *The American Economic Review*, Vol. 56(4), pp. 758-776.

Kydland, F. and Prescott, E., 1982, “Time to Build and Aggregate Fluctuations”, *Econometrica*, Vol. 50(6), pp. 1345-1370.

Lagarde, C., 2020, “The monetary policy strategy review: some preliminary considerations”, speech at the “ECB and Its Watchers XXI” conference, Frankfurt am Main, 30 September 2020.

Mertens, T.M. and Williams, J.C., 2020, “Tying Down the Anchor: Monetary Policy Rules and the Lower Bound on Interest Rates, in *Strategies for Monetary Policy*, Hoover Institution.

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

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

Orphanides, A. and van Norden, S., 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.

Pesek, B., 1963, “Growth, Capacity Output, and the Output Gap”, *The Review of Economics and Statistics*, Vol. 45, No. 3, pp. 294-297.

Plosser, C.I., and Schwert, W., 1979, “Potential GNP: Its Measurement and Significance - A Dissenting Opinion”, *Carnegie-Rochester Conference Series on Public Policy*, pp. 179-186.

Powell, J., 2020, “New Economic Challenges and the Fed's Monetary Policy Review”, Speech delivered at "Navigating the Decade Ahead: Implications for Monetary Policy," an Economic Policy Symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming.

Rabanal, P. and Taheri Sanjani, M., 2015, “Financial Factors: Implications for Output Gaps”, IMF Working Paper 15/153.

Rachel, L. and Summers, L., 2019, “On Falling Neutral Real Rates, Fiscal Policy, and the Risk of Secular Stagnation”, Brookings Papers on Economic Activity.

Romer, D., 2020, “In Praise of Confidence Intervals”, NBER Working Paper No. 26672.

Smets, F., 2002, “Output Gap Uncertainty: Does it Matter for the Taylor Rule?”, Vol. 27 pp. 113-129.

Tereanu, E., Tuladhar, A. and Simone, A., 2014, “Structural Balance Targeting and Output Gap Uncertainty”, IMF Working Paper 14/107.

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.