IEO
Independent Evaluation Office
of the International Monetary Fund

BACKGROUND PAPER

BP/14/01

An Assessment of IMF Medium-Term Forecasts of GDP Growth

Carlos de Resende

Address. 700 19th Street, N.W., Washington D.C. 20431, U.S.A. Telephone. +1 202 623 7312 Fax. +1 202 623 9990 Email. ieo@imf.org Website. www.ieo-imf.org

 BP/14/01

IEO Background Paper
Independent Evaluation Office
of the International Monetary Fund

An Assessment of IMF Medium-Term Forecasts of GDP Growth

Prepared by Carlos de Resende

February 12, 2014

Abstract

This paper assesses the IMF medium-term forecasts—i.e., forecasts for three, four, and five years ahead—of GDP growth published in the *WEO*, covering most IMF member countries, over the period 1990–2012. Evidence is drawn from surveys and interviews with IMF staff, country authorities, and representatives of the private sector, as well as a quantitative analysis of forecast errors. Results indicate that IMF medium-term forecasts of GDP growth meet basic efficiency standards for about 70–80 percent of the member countries. There is no evidence of a systemic or built-in organizational bias specifically caused by the approach used by the IMF to produce medium-term forecasts. Stakeholders have a generally positive view of such forecasts and highly value the Fund's medium-term analysis based on them. On the other hand, the tendency to overpredict GDP growth (i.e., an optimistic bias), previously found in other studies, exists for several countries in all IMF area departments and regardless of development stage and IMF program participation status. The paper argues that more attention should be placed on constructing a unified view about medium-term growth potential in major regions and countries to guide desk economists in their forecasts.

JEL Classification Numbers: C530

Keywords: Medium-term forecast, forecast evaluation, potential output, IMF, *World Economic Outlook*

Author's E-Mail Address: cderesende@imf.org

Contents Page

Tables

Annexes

Abbreviations

AE	advanced economies
AFR	African Department
APD	Asia and Pacific Department
AR	auto-regressive
BN	Beveridge-Nelson
BPF	band-pass filter
CA	current account
DSA	debt sustainability analysis
DSGE	dynamic stochastic general equilibrium
EBA	external balance assessment
EC	European Commission
ECF	Extended Credit Facility
EFF	Extended Fund Facility
EME	emerging market economies
ESAF	Enhanced Structural Adjustment Facility
ESF	Exogenous Shocks Facility
EUR	European Department
FCTA	First Credit Tranche Arrangements
GDP	gross domestic product
GFSR	Global Financial Stability Report
GRA	general resources account
HP	Hodrick-Prescott
LIC	low-income countries
MA	moving average
MCD	Middle East and Central Asia Department
OECD	Organization for Economic Co-operation and Development
PCL	Precautionary Credit Line
PF	production function
PLS	pooled least squares
PRGF	Poverty Reduction and Growth Facility
PRGT	Poverty Reduction and Growth Trust
RBC	real business cycle
REER	real effective exchange rate
RER	real exchange rate
RS	regime switching
SAF	Structural Adjustment Facility
SBA	Stand-by Arrangement
SCF	Standby Credit Facility
SS	state space
SVAR	structural vector auto-regressive
TFP	total factor productivity
WHD	Western Hemisphere Department
WEO	*World Economic Outlook*

I. INTRODUCTION

1. This background study for the IEO evaluation of IMF forecasts complements the literature by focusing on medium-term forecasts (i.e., forecasts for three, four, and five years ahead) of the growth rate of the real Gross Domestic Product (GDP) that is available in the *World Economic Outlook (WEO)* database.[1]

2. Virtually all existing studies that evaluate IMF forecasts focus on short-term forecasts. Although a few (e.g., Faust, 2013)[2] report findings for longer forecast horizons, substantially less space attention has been dedicated to these forecasts, which are used directly or indirectly in many policy-relevant items in the IMF's surveillance and lending activities, and are also regarded as important benchmarks by both country authorities and the private sector for their own analysis and decision making.

3. Systematic errors in medium-term forecasts can significantly affect assessments of the sustainability of fiscal policy frameworks, estimates of equilibrium real exchange rates, measures of sustainable (and desirable) current account positions, and other critical elements of the Fund's analysis, including risk scenarios of the medium and long terms. For that reason, and because of the importance of medium-term forecasts of real GDP growth for the projected path of underlying economic conditions of member countries beyond the temporary effects of the business cycle, a closer look at how these forecasts are conducted at the IMF, and their overall quality, is a necessary part of any evaluation of IMF forecasts.

4. One reason to treat medium-term forecasts separately is that they entail specific methodological challenges, although the forecast process at the IMF integrates both short- and medium-term forecasts in a single framework. The key difference from short-term forecasting is the relative importance placed on cyclical versus structural determinants: the longer the forecasting horizon, the greater the importance attached to structural factors. Hence forecasters of medium-term developments must try to identify trends that capture structural, as opposed to cyclical,[3] forces in the variables being forecasted.[4]

---

[1] For the main report of the IEO evaluation, see IEO (2014a).

[2] Faust (2013) refers to an ongoing external evaluation of IMF forecasts, commissioned by the IMF. The final report is not yet available. The author had access to a preliminary version (February 2013).

[3] For this reason, methods typically used to forecast GDP growth in the short-term tend to focus on the informative content of recently past GDP growth performance and/or factors that are sensitive to the business cycle. See Chauvet and Potter (2012) for a survey of the recent literature on methods used for forecasting output growth in the short term.

[4] In the case of GDP, for example, that trend is often associated with the long-run aggregate supply curve, usually interpreted as the level of *potential output*. See Congdon (2008) and Section III below.

5. The study has three broad objectives: (i) to illustrate the importance of accurate and unbiased medium-term forecasts of GDP growth in the IMF's role of providing policy advice to member countries; (ii) to better inform stakeholders on the existing methods to estimate potential GDP growth, which is a concept critical to forecasts of actual GDP growth over longer horizons, and, in doing so, discuss the strengths and pitfalls of such methods, as well as their use by IMF staff; and (iii) to assess the overall quality of IMF medium-term forecasts of GDP growth from the standpoint of informational efficiency and accuracy, while looking at how these forecasts are perceived by potential users among country authorities and in the private sector.

A. Scope of the Analysis

6. The study focuses on forecasts of real GDP growth, as this is judged the most important variable according to a survey of country authorities and subsequent follow-up interviews.[5] Concentrating on GDP, as a key factor underlying IMF forecasts of both fiscal and current account balances, also helps the report effectively achieve a broad coverage without becoming excessively taxonomic and lengthy. (Henceforth, the forecasts of real GDP growth will be simply referred to as "medium-term forecasts," unless explicitly stated otherwise.)

7. For a number of reasons we restrict the quantitative analysis to forecasts made for the *World Economic Outlook* (*WEO*) rather than forecasts published in Article IV reports. First, *WEO* forecasts are more frequent and are issued at regular intervals (twice a year at roughly at the same dates for the entire IMF membership). This facilitates their comparison with those of other agencies that provide forecasts on a regular basis for many economies. Second, *WEO* forecasts have been analyzed in commissioned studies of IMF forecasting performance since the 1980s, allowing us to make comparisons with those studies and to assess how the IMF learns from its past forecasting performance. Third, the *WEO* data are readily available, being organized in a comprehensive dataset, while no unified dataset for Article IV forecasts seems to be available. Moreover, as explained in the main evaluation report (IEO, 2014), except for reasons related to timing there should be no substantial differences between *WEO* and Article IV forecasts, since their preparation are integrated in the same general process used in IMF forecasts.[6]

---

[5] See Genberg and Martinez (2014a) for details on the survey, which was undertaken for the IMF Forecasts evaluation by IEO. The survey results show that, among the forecasts in the *WEO*, those of GDP growth are unambiguously the most valued by both country officials and representatives from the private sector. In telephone interviews with country authorities, virtually all participants clearly mentioned GDP growth as the most important variable. Other variables such as inflation and measures of fiscal balance and public debt have been covered in earlier commissioned evaluations of IMF forecasts and are discussed in Luna (2014), which is also part of this IEO evaluation.

[6] The process is described in Genberg, Martinez, and Salemi (2014).

8. The analysis concentrates on point forecasts. Clearly, informed views about the future require a broader perspective than point forecasts by themselves. But point forecasts can nonetheless be viewed as the basis, or starting point, for broader sets of considerations about future economic developments. Any problems with the former are likely to distort the latter.[7]

9. Evidence for the analysis is drawn from (i) the opinion survey conducted by the IEO evaluation team with member country authorities, market participants from the private sector, and IMF desk economists on various aspects of IMF forecasts, including medium-term forecasts; (ii) post-survey follow-up interviews with 50 randomly selected IMF economists, which were aimed at helping the evaluation team refine and qualify its understanding of the survey results; (iii) bilateral interviews with selected members of IMF staff, the Board of Executive Directors, representatives of the private sector, and country officials;[8] and (iv) the author's own statistical analysis of three-, four-, and five-year-ahead IMF forecasts of GDP growth available in the *WEO* database.

B. Outline of the Paper

10. Section II explains why medium-term forecasts are important to the technical quality of the IMF's work. It then illustrates how biases in these forecasts can distort the understanding of medium-term risks and the policy advice that can follow from them. Section III describes the different methods in common use for estimating potential output, which is a key variable underlying medium-term forecasting; technical details of these methods are provided in Annex 1. Section IV describes the Fund's medium-term forecasting process and methods and discusses some issues posed by the current arrangements. Section V presents the findings of statistical analysis of the accuracy and efficiency of the forecasts, covering the period 1990–2012; supporting details of the methodology and the findings are given in Annexes 2 and 3. Section VI reports how the forecasts are perceived by users—member country authorities and representatives of the private sector—and by staff; further details are given in Annex 4. Each of Sections II through VI concludes with a short summary of its main findings. Section VII concludes the paper with an overall summary and discussion of findings, followed by some recommendations for improving the IMF's medium-term forecasting process.

---

[7] As described later in this paper, risk scenarios and the analysis of driving forces behind the path of the variables forecast are also important; they are highly valued by country officials according to the evaluation survey, and are being increasingly incorporated in IMF flagship documents and Article IV reports, likely in response to the recommendations from the commissioned external evaluation of IMF forecasts by Timmermann (2006). The paper, however, does not directly evaluate the quality of such broader analysis of future developments, as it entails a different set of methods and objectives relative to pure forecasting.

[8] This includes about 20 telephone interviews with country officials.

II. WHY ARE ACCURATE MEDIUM-TERM FORECASTS IMPORTANT?

11. IMF medium-term forecasts—for three, four, and five years ahead—of the growth rate of real GDP in the *WEO* database may be interpreted as the views of the staff about the sustainable trajectory of member countries' economies over the foreseeable future. This sustainable path encompasses the staff's assessment of the trend associated with the long-run aggregate supply curve, usually interpreted as a measure of *potential output*, to which the actual real GDP level reverts as the transitory effects of macroeconomic disturbances associated with the business cycle dissipate, and that is consistent with a stable inflation rate and "normal" rates of unemployment and overall economic slack. Medium-term forecasts are crucial for several policy-relevant activities and products in the IMF's surveillance and lending agendas. They are a key underlying feature of the Fund's advice to individual member countries and of its assessment of risks in the global economy.

12. Maintaining a good quality of forecasts is an important aspect of the Fund's overall reputation. The IEO survey results indicate that country authorities and the private sector have a generally positive view of IMF forecasts, both short- and medium-term, even though the literature has occasionally reported biases and informational inefficiencies in these forecasts for some member countries.[9] The survey results also show that IMF medium-term growth forecasts are often viewed as an important benchmark by both country authorities and the private sector. These stakeholders often use IMF forecasts as inputs for their own analysis and decision making. Indeed, as detailed in Section VI below, more than half of the survey respondents from the private sector think that for policy making purposes medium-term forecasts are more important than short-term forecasts.[10]

13. Subsection A below describes the role of medium-term forecasts in some key IMF activities and illustrates how forecast errors can adversely affect policy making and advice. Subsection B describes how difficulties in estimating potential GDP—a variable that can never be observed but that nonetheless plays a crucial role in medium-term forecasting—translate into uncertainties and errors in medium-term forecasts. The illustration emphasizes the need for a careful choice and application of methods, and also the need to ensure that users appreciate the uncertainties surrounding the real growth forecasts that emerge. Subsection C briefly discusses the potential importance of longer term forecasts and estimates of potential output and the output gap in short-term projections. Subsection D summarizes the section's main points.

[9] See for example Timmermann (2006), Freedman (2014), Faust (2013), and Genberg and Martinez (2014b).

[10] See also Genberg and Martinez (2014a).

A. Use of Medium-Term Forecasts in Key IMF Products

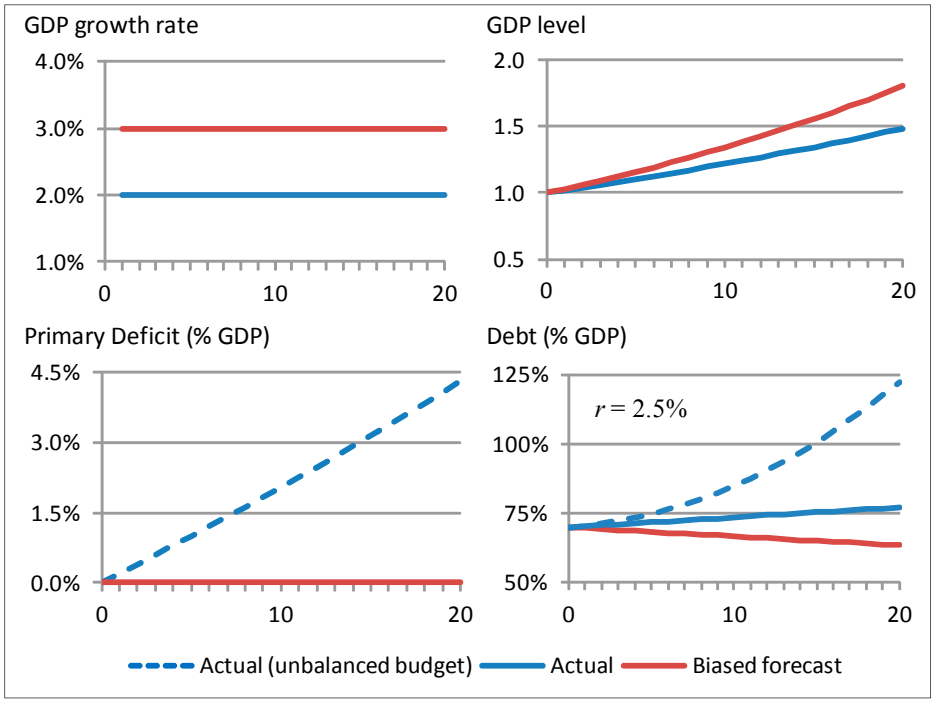
14. This subsection illustrates how medium-term forecasts affect the conclusions drawn in IMF analyses of country situations and global economic trends, and hence the advice the Fund offers to its members.

Debt Sustainability Analysis (DSA)

15. Debt sustainability analysis (projecting the evolution of a country's government debt, relative to its underlying ability to generate income) is an activity in which medium-term forecasts of GDP growth play a key role. Erroneous forecasts of growth may produce a distorted view of the future debt level and lead to misguided policy advice today. IMF staff research (IMF, 2004) suggests that poorly made projections of real GDP growth for longer horizons may greatly undermine assessments of debt sustainability.

16. As an illustration, consider the hypothetical economy represented in Figure 1, showing the level and growth rate of GDP, as well as the primary deficit and government debt as shares of GDP. The blue lines in all panels represent the expected path of the variables when forecasts are unbiased (i.e., the actual out-turns). It is assumed that the (true) sustainable annual GDP growth rate is $g = 2$ percent, but that its forecast is optimistically biased, at $\hat{g} = 3$ percent (top-left panel)—which implies a path for the level of GDP consistently above the true level (top-right panel).

Figure 1. Effect of a 1 percentage point optimistic bias in forecast of GDP growth



Source: Author's simulations.

17. Consider an annual real interest rate of r = 2.5 percent, an initial debt-to-GDP ratio of 70 percent, an initial balanced (primary) fiscal position, and a fixed tax rate that keeps tax revenues at 20 percent of GDP. In addition, assume that government spending is exogenously set to keep the primary budget balanced. The debt-to-GDP ratio, d_t, evolves according to:

$$d_t = \frac{(1+r)}{(1+g)} d_{t-1} + def_t,$$

where def_t is the primary deficit as share of GDP.

18. First, assume that the authorities succeed in keeping the budget balanced over the foreseeable future, such that def_t = 0 at all times. Under the unbiased forecast, the debt-to-GDP ratio has an upward trend because $r > g$ (that is, debt grows faster than GDP). The biased forecast, on the other hand, suggests a decreasing path, since $r < \hat{g}$. By year 20, the bias in the GDP growth forecast would imply a 63 percent debt-to-GDP ratio, about 14 percentage points below the true level, as indicated by the solid blue and red lines in the bottom-right panel of Figure 1. This difference is due purely to the effect of the overprediction of GDP growth: the same debt profile is being normalized by different levels of GDP.

19. What if the optimistic bias in GDP growth forecasts also affects the evolution of debt itself? Assume that the authorities rely on forecasts of GDP growth to set the level of government spending according to the expected revenue from taxes. Again, their goal is to keep the budget balanced, but biased forecasts may mislead them in their expectations about the evolution of public finances. Given the fixed tax rate and the true sustainable growth rate, tax revenues actually grow at 2 percent per year, although the authorities expect 3 percent if they are relying on the biased forecast.

20. If, at time 0, actual spending is allowed to grow at that same rate to keep up with the forecast path of tax revenues, under the biased forecast the budget is expected to be balanced but in reality the primary deficit as a share of GDP increases to about 1 percent of GDP by year 5 and to 4.5 percent by year 20 (dashed blue line in the bottom-left panel of Figure 1). The path of the debt-to-GDP ratio would actually be explosive, rather than declining as under the biased forecast (bottom-right panel), and the difference between the two paths would amount to 9 percentage points by year 10 and 37 percentage points by year 20. Although this is an extreme case, it helps to illustrate the importance of accurate and unbiased medium-term forecasts for the analysis of debt sustainability.[11]

[11] The exercise assumes that no correction in government spending is made over the years as expectations about growth are not realized. A more realistic scenario perhaps lies between the two cases discussed above. In the context of IMF programs, IMF (2004) reports a similarly important effect of optimistic biases in medium-term forecasts on the DSA.

21. The Fund's new DSA framework for market access countries (IMF, 2013b) requires staff to look at GDP forecast errors (albeit not necessarily in the medium to longer term), indicating that the IMF acknowledges that forecast errors can affect its assessment.

Pilot External Balance Assessment (EBA)

22. The pilot EBA is another IMF product that uses medium-forecasts. It provides assessments of whether current account balances and exchange-rate levels are consistent with fundamentals and "appropriate" policies for a group of 49 advanced and emerging market economies, which together account for about 90 percent of global GDP. As shown below, biases in the growth forecast for a country or for its trading partners can change the assessment of the norm for the real effective exchange rate, potentially leading to situations in which the rate can be judged under- or overvalued relative to the norm, when it is actually in equilibrium.

23. The EBA methodology (described in IMF, 2013a) relies specifically on five-year-ahead forecasts of GDP growth from the *WEO* database. It starts with a descriptive phase, where a panel regression is estimated and used to uncover which variables are important to determine current accounts and real exchange rates across countries, and then uses the results in normative evaluations. The latter consist of identifying "policy gaps"—differences between the current levels of some policy variables used in the panel regression and their desired levels, or "norms," as identified by desk economists. The normative analysis thus depends on the regression results, which in turn use data from the *WEO* forecasts.

24. One of the variables used in the EBA regressions is the difference between a specific country's five-year-ahead forecast of GDP growth and the cross-country weighted average. For instance, the EBA regression for the real effective exchange rate (REER) of country *i* is:

$$log(REER_i) = \hat{\beta}X + \hat{\gamma}P + 2.35\left[\hat{g}_{t,t+5}(i) - \hat{g}_{t,t+5}(w)\right] + u,$$

where $\hat{g}_{t,t+5}(i)$ and $\hat{g}_{t,t+5}(w)$ are the five-year-ahead forecasts of GDP growth (made at year *t*) for country *i*, and for the world economy (computed as a weighted average of individual countries' forecasts); X is a vector of variables that affect the REER through estimated coefficient $\hat{\beta}$ but are not fully under the country authorities' control; P is the set of policy variables (e.g., fiscal variables) that affect the REER according to estimated coefficient $\hat{\gamma}$ and are typically under the discretion of authorities; and u is the regression residual.

25. Denoting the difference in forecasts by $Q = \hat{g}_{t,t+5}(i) - \hat{g}_{t,t+5}(w)$, and the set of desired policies by P^*, the gap between the actual value of the REER and its "norm," or desired level, is given by:

$$REER\ gap = \underbrace{log(REER_i)}_{\text{actual}} - \underbrace{\left(\hat{\beta}X + \hat{\gamma}P^* + 2.35Q\right)}_{\text{"norm"}} = \underbrace{\hat{\gamma}(P - P^*)}_{\text{effect of policy gaps}} + u.$$

26. Note that, everything else constant, a measurement error in Q due to biases in either $\hat{g}_{t,t+5}(i)$ or $\hat{g}_{t,t+5}(w)$ will distort the assessment of the REER norm and gap. For a given value of the policy gap, this distortion due to the bias is accommodated in the residual series.

27. According to the EBA equation, an increase of one percentage point in country *i*'s expected growth rate, relative to the world growth rate, induces an appreciation of 2.35 percent in that country's REER.[12] Assume that the country forecast $\hat{g}_{t,t+5}(i)$ is measured with an optimistic bias that is larger than the (average) optimistic bias observed in $\hat{g}_{t,t+5}(w)$. In that case, an unbiased country forecast would produce a lower predicted value for the REER norm than would a biased forecast. The same would happen if the pessimistic bias in the world (average) forecast $\hat{g}_{t,t+5}(w)$ is larger than in $\hat{g}_{t,t+5}(i)$. That is, optimistic (pessimistic) biases in country (world) forecasts, aside from their potential effect on the estimated results themselves, induce the EBA methodology to predict (i) an excessive appreciation of the norm REER and (ii) a gap between the current REER and its norm that is narrower than the true gap.

28. Assume that the true REER gap is zero, but wrongly assessed because of an optimistic bias in country five-year forecasts of GDP growth. The policy advice based on the EBA could thus mislead the country authorities into letting the REER appreciate when the latter is actually in line with the desired policies and fundamentals. Similarly, even if there is no bias in the forecasts for a given country, optimistic biases in GDP growth forecasts for other economies could induce a call for exchange-rate depreciation when none is needed in order to close the REER gap. Clearly, the specific distortion just discussed is not caused by shortcomings in the EBA methodology but is simply the result of biased forecasts.

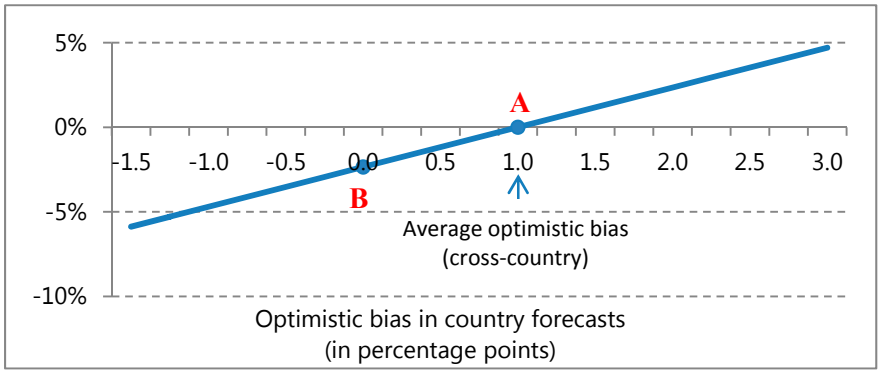
29. Figure 2 shows how this distortion changes depending on the bias in country forecasts, $\hat{g}_{t,t+5}(i)$, relative to the bias in world forecasts, $\hat{g}_{t,t+5}(w)$. Positive numbers on the horizontal axis indicate optimistic biases. Consider an optimistic bias in world GDP growth forecasts of 1 percentage point.[13] In the particular case that both forecasts show the same bias (point **A**, in Figure 2), the EBA methodology does not produce any additional overvaluation of the REER norm. In general, however, these optimistic biases do not cancel each other out. To the right of point **A**, country forecasts are increasingly more optimistic relative to average forecasts, inducing an overvaluation of about 5 percent when the degree of optimism in country forecasts is close to three times the cross-country average optimistic bias. To the left of point **A**, country forecasts are less optimistic than the world average and eventually

[12] Fast growing economies tend to have deficits in their current accounts, consistent with the appreciation of the REER.

[13] If θ is the share of country *i*'s in world GDP, then the total effect of biases in forecasts of GDP growth on the assessment of the REER is equal to $2.35(1-\theta)$ times the country bias. By normalizing the bias in world GDP growth forecasts to 1 percentage point, we abstract from the indirect effect of the bias in country forecasts through the average bias.

become negative, or increasingly pessimistic. Even with no bias in country forecasts (point **B**), the 1 percentage point optimistic bias in the cross-country average forecast leads to an undervaluation of the REER norm of about -2.4 percent, inducing the EBA methodology to call for a depreciation of the exchange rate.[14]

Figure 2. EBA methodology: additional appreciation of the norm REER caused by optimistic biases in five-year-ahead forecasts of GDP growth



Source: Author's simulations.

Risk analysis in medium-term scenarios

30. Because forecasts involve uncertainty, point forecasts need to be complemented by a broader set of analytical tools. And because uncertainty typically grows as the forecast horizon lengthens, such complementary tools are especially important in the case of medium-term forecasts. Increasingly, as shown below, the IMF's flagship and bilateral surveillance reports discuss future developments in terms of alternative scenarios and risks around baseline scenarios, which usually take the point, or central, forecasts as inputs, or benchmarks.

31. Survey-based evidence shows that country officials highly value the Fund's analysis of medium-term scenarios for the global economy, specific regions, and individual economies built on alternative assumptions to those underlying the point forecasts.[15]

32. Medium-term scenarios and risk analyses increasingly feature in IMF's flagship and bilateral surveillance reports (Table 1).[16] IMF products—including the flagship publications *WEO*, *GFSR*, and *Fiscal Monitor* (FM), as well as spillover reports, external sector reports,

[14] A similar distortion will occur in the assessment of current account balances according to the EBA, which relies on a similar methodology. In the current-account regression, the coefficient of Q, the difference in the five-year-ahead forecast of GDP growth relative to its cross-country weighted average, is about -0.5.

[15] See below and, also, Genberg and Martinez (2014a).

[16] Some IMF historians (see Boughton, 2001, p. 227) consider these medium-term scenarios to have become even more important than short-term forecasts.

and early-warning and vulnerability exercises—increasingly present analyses of medium-term issues, often relying on risk scenarios around baseline projections that derive from point forecasts. Annex 5 catalogs the medium-term analyses in these products.

Table 1. Medium-term analysis in IMF products

IMF product	No. of reports	Reports with fan charts	Fan charts per report (average)	Reports with figures or tables of medium-term scenarios	Figures or tables of scenarios per report (average)	Reports with sections about medium-term issues [1]	Sections about medium-term issues per report (average)
				2000–2013			
World Economic Outlook	28	57%	0.7	93%	2.8	64%	1.3
Fiscal Monitor[2]	10	20%	0.3	60%	1.0	60%	1.4
Global Financial Stability Report	26	8%	0.1	19%	0.2	19%	0.3
Article IV Reports [3]	56	4%	0.04	96%	4.6	84%	1.3
				2006–2013			
World Economic Outlook	16	100%	1.3	94%	2.7	50%	0.9
Fiscal Monitor [2]	10	20%	0.3	60%	1.0	60%	1.4
Global Financial Stability Report	16	13%	0.1	31%	0.3	13%	0.1

Source: Author's calculations.

(1) Based on search for selected key-phrases on section titles.
(2) The Fiscal Monitor started in 2009.
(3) Most recently available reports for 56 randomly selected countries.

33. Of the 28 editions of the *WEO* published since 2000, almost two-thirds have sections or chapters with titles containing key-phrases that suggest medium/long-term subjects17, while 57 percent use "fan charts" to describe the uncertainty around medium-term central forecasts, and 93 percent contain figures, charts, or tables describing medium-term scenarios. On average, a typical WEO report contains about three figures or tables with scenarios and at least one section about medium-term issues. The use of fan charts to describe risks to the baseline scenario has risen dramatically since 2006; all WEO editions published since then have contained at least one fan chart.

34. Article IV reports use charts and tables containing medium-term scenarios more frequently than the *WEO*, although their use of fan charts is relatively rare (only 4 percent of these reports). Within a random sample of 56 Article IV reports, selected from the most recently available for the entire IMF membership, 96 percent use such charts and tables and the number of such tools per report is more than twice that in the *WEO*.

35. Because scenarios and risk assessments for the global and individual economies are constructed using central or point forecasts of GDP growth as their baseline, they are highly

[17] The key-phrases searched by the evaluation team, which do not exhaust all possibilities, are "medium term," "long term," "potential output," "output gap," "sustainable growth," "structural reform," "structural change," and "debt sustainability."

sensitive to the quality of these forecasts. Systematic errors in point forecasts can have significant effects on these scenarios and compromise the risk assessments, potentially distorting understanding of the risks to financial and macroeconomic stability and hampering the quality of the Fund's policy advice. As the reliance on these assessments grows, so does the risk that problems in point forecasts will affect the policy advice provided to member countries.

36. Three examples are presented in Figure 3, panels (a)-(c). In all panels, the red dashed line shows the average forecast for the *level* of real GDP implied by the biased forecasts of GDP growth, going from forecasts for the current year to five years ahead. The blue dashed lines represent the level of GDP implied by the observed average growth rate, which also represents the expected path for GDP under an unbiased forecast. The solid red and blue lines represent 90 percent confidence intervals around the biased and unbiased forecasts for the level of GDP, respectively.[18] Therefore, the red and blue dashed lines represent point forecasts that are the mid-point of the confidence intervals, which convey the uncertainty associated with the central forecasts.

37. Both point forecasts and analysis based on risk assessments and scenarios are generally highly valued by country authorities. In assessing the value added by point forecasts to support policy decision making in their own country and in relation to their understanding of the world economy, between 73 percent and 91 percent of the country officials who responded to the survey said they attach "high value" or "very high value" to point forecasts, while between 11 percent and 25 percent attached "low value" or "no value."[19] These figures compare to between 81 percent and 91 percent of country officials who highly value scenarios, risk assessments, and analyses of regional and global economic prospects.[20]

---

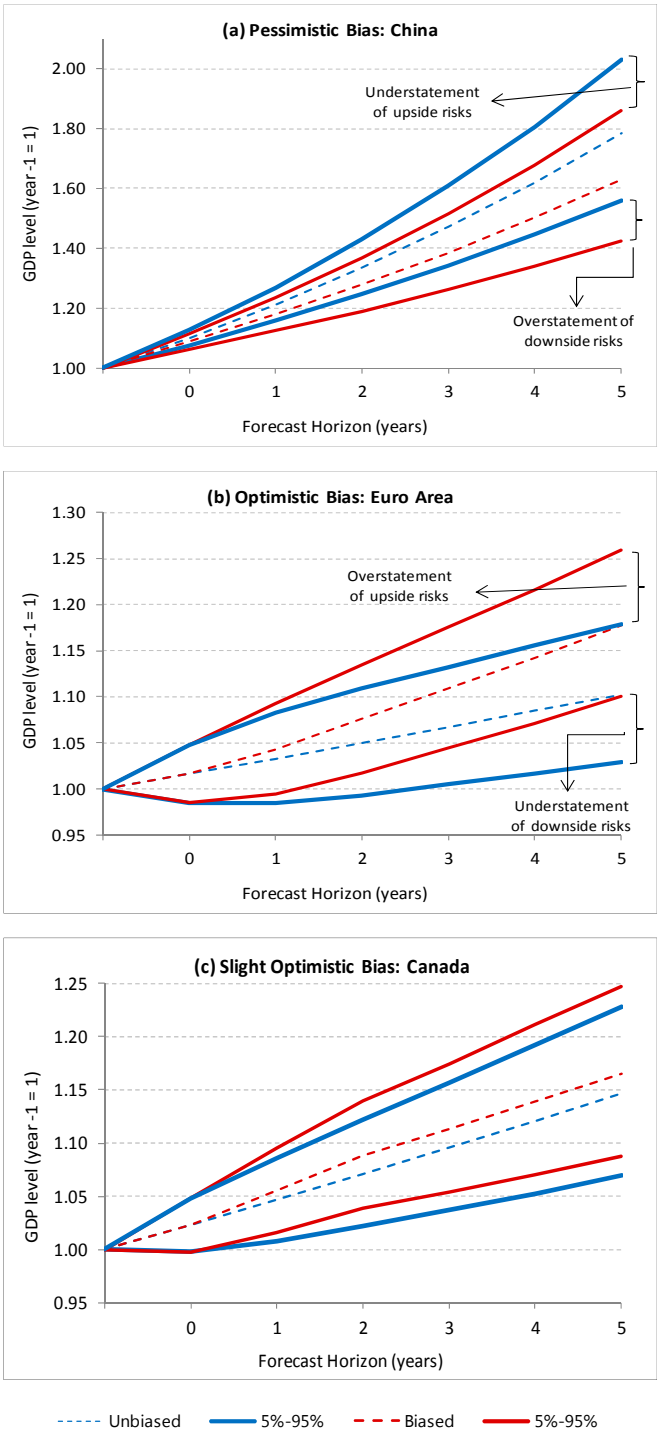
[18] The unbiased profiles of the level of GDP (blue dashed lines) were computed by simply accumulating the observed average GDP growth rate as reported in the most recent *WEO* (October 2013). The biased profiles (red dashed lines) were constructed by subtracting the average forecast error for each horizon—using all Spring vintages of the *WEO* from 1990 to 2012, as reported in Annex 2 of the main report (IEO, 2014a)—from the unbiased path. Without the models or approaches that generated the forecasts—which may have changed over successive rounds of forecasts in different vintages of the *WEO* or may not even have a statistical/econometric representation—the probability distributions associated with the forecasts are not available to the author and, therefore, true confidence intervals cannot be constructed. To construct a proxy for these confidence intervals, a normal distribution with standard deviation obtained from the series of forecasts across *WEO* vintages, for each forecast horizon, was used. The same dispersion around the central forecast was applied to both the unbiased and biased forecasts, generating the confidence intervals represented by the solid blue and red lines, respectively.

[19] Respondents assessed the value of point forecasts in the *WEO* and in Article IV reports for the analysis of their own country, neighboring countries, their own region, and the world, as well as for emerging and advanced economies as separate groups. See Genberg and Martinez (2014a).

[20] See Genberg and Martinez (2014a), pp. 21–23. Officials tend not to value forecasts for their own country as highly, because they consider themselves better informed than the IMF. Thus when the questions about their own country and about forecasts in Article IV reports are excluded, the high value attached to point forecasts is

(continued…)

Figure 3. Effect of biased forecasts on risk analyses



Source: Author's calculations using *WEO*.

even more evident and becomes essentially identical to the value attached to other aspects of forecasts that are presumably important for medium-term analysis, such as risk assessments and scenarios.

38. Panel (a) shows an example of a member country (China) for which the evaluation team found evidence of a statistically significant pessimistic bias in IMF forecasts of GDP growth—that is, projected GDP growth has been systematically below out-turns. Panels (b) and (c) show cases of relatively large (Euro Area) and small (Canada) optimistic biases, respectively.

39. Biases in point forecasts imply a shift in the confidence interval relative to that associated with unbiased forecasts. With pessimistic biases (panel a), the confidence interval is shifted downwards, leading to the understatement of upside risks and the overstatement of downside risks: the estimated probability of realization of scenarios associated with higher growth rates becomes lower than it should be under unbiased forecasts, while the probability of scenarios of lower growth becomes higher. The upper limit of the confidence interval using the biased forecasts (red solid line) barely exceeds the mid-point of the interval constructed using the unbiased forecasts (blue dashed line).

40. When forecasts are biased towards overoptimism, the opposite occurs: the probability distribution of risks is shifted up, as in Figure 3 panels (b)-(c). As seen in Section V below, such optimistic biases are found to be the most common type of bias in forecasts across the IMF membership. When optimistic biases are not large (panel c), a medium-term risk analysis constructed using the biased forecasts is not much affected by the errors. Note how the biased forecast (red dashed line) still falls well inside the "right" confidence interval (blue solid line).

41. However, if biases are large enough relative to the level of uncertainty measured by the width of the confidence interval, as in panel (b), then the baseline scenario based on the biased forecasts may approach the upper bound of the appropriate confidence interval, while the baseline scenario based on unbiased forecasts (blue dashed line) will touch the lower bound of the "wrong" confidence interval (red solid line). This means that high-growth scenarios that are very unlikely to materialize will be included in the confidence interval, while a large portion of the downside risks will be completely ignored by the analysis.

B. Potential Output and Output Gap

42. Making medium-term forecasts is inherently a risky and uncertain business. In particular, weaknesses in the analytical underpinnings of these forecasts can lead to inaccurate forecasts and hence misguided policy advice. Here we illustrate the possible consequences of making errors in the estimation of potential GDP and the output gap, which (as discussed in Section III below) are key concepts underlying much of the Fund's medium-term forecasting.

43. Potential output is usually interpreted as the sustainable level of GDP both in the physical sense (given the country's available resources) and in the economic sense (consistent with a stable rate of inflation)—and represents the level to which the economy reverts when the effects of temporary shocks that cause cyclical fluctuations dissipate. According to this interpretation, estimates of potential output should help identify the pace of sustainable

growth of GDP over the long term. The related notion of *output gap*—the difference between the measured level of GDP and potential output—is a key indicator of the degree of slack in the economy. The measure of the output gap is typically used in short-term forecasts of inflation and the measurement of cyclically adjusted fiscal and current account balances.

44. During interviews conducted by the evaluation team with IMF staff, it became clear that having an estimate of the level of potential output is an important step in the process of obtaining medium-term forecasts of GDP growth. A large majority of the interviewees explained that the medium-term outlook (not only for GDP, but also other variables) is anchored on expectations of potential output, which become the cornerstone for medium-term forecasts of GDP growth. Indeed, references to potential output and the output gap appear in about 30 percent of all staff memoranda (SM) sent to the IMF Board since the year 2000.[21]

45. However, potential GDP is an unobserved variable, which makes its estimates particularly problematic because they cannot be checked against actual outturns even ex post. Methods available for estimating potential GDP have several shortcomings and involve a large degree of uncertainty. One problem associated with virtually all of them is that the results are sensitive to updating the sample on which they are based.[22] Estimates of *past* potential growth rates will in general change when new data become available, implying considerable uncertainty in the original estimates. This is illustrated in the figures that follow, which plot the growth rate of potential GDP as recorded in different *WEO* vintages for the United States (Figure 4) and for a group of 16 developed economies (Figure 5).[23]

46. Relative to the growth rates of potential GDP published in the *WEO* Spring 2012 (red line), the estimates for any year in the sample substantially change across different *WEO* vintages. For example, for the United States, the *WEO* of Spring 2012 estimates the growth rate of potential GDP in 1998 as 3.5 percent, but the estimates for that same year have been as low as 2.3 percent (*WEO* Spring 1998) and as high as 4.3 percent (*WEO* Spring 2010), a difference of 2 percentage points.[24]

---

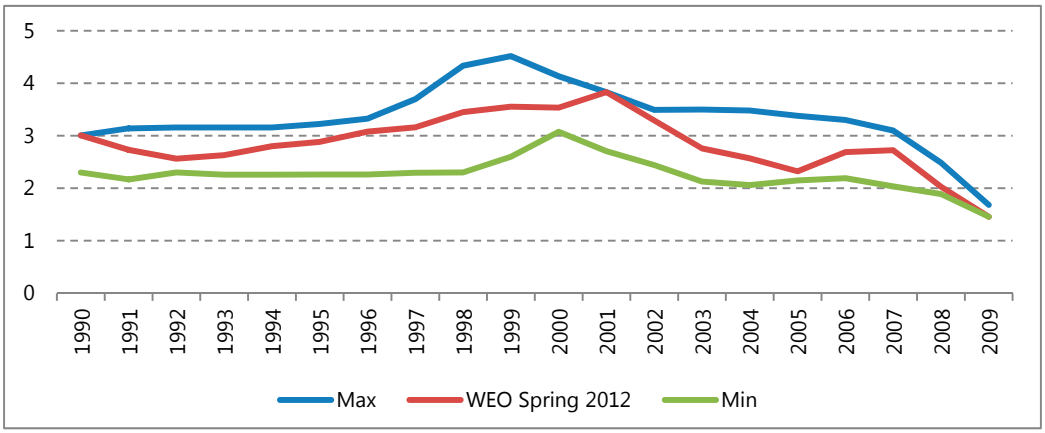
[21] Considering the universe of SMs in Cyberdocs (a documentation database available to IMF staff) , this is the result of a search of "potential output" or "potential growth" or "potential GDP" or "long term growth" or "sustainable growth" or "output gap."

[22] This is the case particularly at the end of sample period, when such estimates are the most important for forecasts and policy analysis.

[23] The 16 economies are: Australia, Austria, Belgium, Canada, France, Germany, Italy, Japan, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, and the United States.

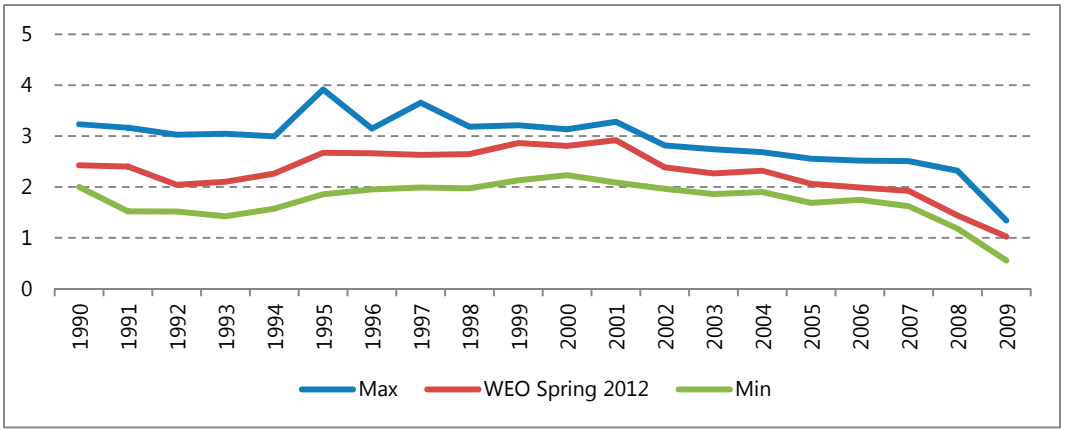
[24] The narrower difference between minimal and maximal estimates of the US potential growth rate at the end of the sample period in Figure 4 does not necessarily mean more precise estimates, but rather reflects the smaller number of *WEO* vintages used to compute the mean, minimal, and maximal potential growth rates for more recent years.

Figure 4. Growth rate of potential GDP in the United States across different *WEO* vintages
(In percent)



Source: Author's calculations using *WEO*.

Figure 5. Cross-country (simple) average growth rates of potential GDP across different *WEO* vintages in 16 OECD economies
(In percent)



Source: Author's calculations using *WEO*.

47. This wide variation in estimates of past growth rates of potential GDP is also seen in other economies and it may be interpreted as a measure of how uncertain these estimates are. It indicates, for example, that forecast errors of the same order of magnitude as the 1 percentage point optimistic bias used above to illustrate the effect of biases in medium-term forecasts on the debt sustainability analysis and external balance assessments are not at all uncommon.

48. The revisions made in the IMF's estimates of potential GDP are partly explained by revisions in the data on past actual GDP.[25] They may also be due to the continuous learning that takes place about the economy's past as new data become available (especially because that potential GDP is an unobservable variable), or to changes in the method being used.

49. However, some of the revisions of potential GDP may reflect IMF forecasters' widespread use of "mechanical" filtering techniques, such as the popular Hodrick-Prescott (HP) Filter, whose results are very sensitive to changes in the sample. As discussed in Section III and Annex 1, some of these techniques can produce estimates of past potential GDP that change dramatically as new observations are included in the analysis.

50. Because estimates of the output gap are conditional on estimates of potential GDP, the uncertainty in the estimates of potential GDP carries over to the estimates of the output gap (recall that the latter is the difference between the measured level of output and potential output). For example, the *WEO* estimate of the output gap for the United States in the year 2000, as a percentage of potential output, was roughly 1.3 percent and 1.9 percent in the Spring and Fall vintages of the *WEO* in that year, respectively, indicating an economy moderately overheated. Later estimates for that same year more than doubled (4.4 percent) in the Fall of 2006, decreased to 0.9 percent (Fall 2009), and went up again to 4.4 percent (Fall 2012). The most recent estimate (Fall 2013) puts the output gap in the United States at about 2.8 percent in the year 2000. These awkward revisions of past estimates indicate how uncertain the estimates of the cyclical state of the economy are and how risky the policy advice based on them can be, despite the fact that the most recent *WEO* vintage may be based on the largest possible information set. Mistakes in the estimation of potential GDP and the output gap can lead to misguided policy advice regarding the appropriate responses of monetary, fiscal, and exchange rate policies to the current cyclical state of the economy.[26]

C. Use of Potential Output and Medium-term Forecasts in Short-term Forecasting and Policy Advice

51. To the extent that forecasts of GDP growth for longer horizons contain information about the evolution of the "normal" path of the economy around which short-term fluctuations will concentrate, they may help boost the accuracy of short-term forecasts.

52. Estimates of potential output are thought to contain information about the economy's supply side and the maximum level of output that can be generated with the full utilization of production factors without additional inflationary pressures, a standard interpretation in

[25] See Genberg and Martinez (2014b) for a discussion of the effect of data revisions on the assessment of forecast errors.

[26] For instance, Orphanides and Wieland (2012) document that problems with the measure of output gap can have deleterious effects on optimal monetary policy analysis.

macroeconomics since Okun (1962). According to this interpretation, estimates of potential output should therefore help identify the pace of sustainable growth of GDP over the long term. This information is not only an important benchmark for medium-term forecasts, but it can also be used in the short-term analysis to assess, for example, inflationary pressures associated with the economy's cyclical position.

53. For instance, a positive output gap indicates that actual GDP is above its supply-side potential, sustainable, non-inflationary level, which may suggest that the economy faces inflationary pressures stemming from excess demand in both goods and labor markets. Short-term forecasts of inflation, the measurement of cyclically adjusted fiscal and current account balances, as well as the policy advice in these situations—to put in place counter-cyclical measures of monetary, fiscal, and exchange rate policies aiming at mitigating the excess demand—critically depend on the measure of the output gap and, thus, on the estimates of potential GDP.

54. Faust (2013) suggests that, when the structure of the economy is changing, it may be better to focus on forecasting the new typical path the economy should follow after the structural changes mature, rather than trying to predict the short-term variability around and leading to that path. Under this notion, Faust shows that private sector long-term forecasts (six- to ten-year ahead forecasts) of GDP growth published by Consensus Economics can be used to improve *WEO* short-term forecasts.[27] At the very least, they reduce the variability of short-term forecasts without major, if any, losses in accuracy.

D. Summary of Findings on the Importance of Accurate Medium-Term Forecasts

55. To sum up the discussion in this section:

- Country authorities highly value the IMF's forecasts, both central (or point) forecasts and analyses based on risk assessments and scenarios, which are becoming increasingly important in IMF products.

- Medium-term forecasts are important for economic planning purposes and are used by the IMF in policy-relevant analytical frameworks such as Debt Sustainability Analysis (DSA) and the Pilot External Balance Assessment (EBA) to inform its surveillance and lending activities.

- Biased medium-term forecasts of GDP growth may distort the results of DSA and EBA, and may lead to major miscalculation of the risks associated with alternative scenarios around the baseline used in medium-term analysis. These biases in forecasts can therefore be reflected in misleading policy advice to member countries.

[27] Consensus forecast are published by Consensus Economics, Inc. See www.consensuseconomics.com.

- The IMF's estimates of the potential GDP for any given year can change dramatically over different *WEO* vintages, even for years past. This great variation is an indication of the difficulty of estimating an unobservable variable and suggests substantial uncertainty around these estimates.

- As a byproduct of the challenges in the estimation of potential GDP, the output gap also becomes very uncertain, affecting the measure of the cyclical position of the economy, which is important for short-term forecasts and policy advice in real time.

56. In light of the challenges to obtain reliable estimates of potential output—a key concept in medium-term forecasts of GDP growth—the next section discusses some of the available methods used IMF to estimate potential output.

III. Methods of Estimating Potential Output

57. Medium- and longer-term forecasts of GDP growth tend to abstract from cyclical factors, which are very hard to predict over longer horizons, and to concentrate on identifying the trend, or "norm," which is usually interpreted as the level of potential output in the economy. Separating the level of actual GDP into a long-run trend and transitory fluctuations around the trend provides a useful notion of the business cycle that is widely used in economic theory, forecasting, and policymaking.[28]

58. In interviews conducted by the evaluation team, IMF desk economists made it clear that for them the estimation of potential output is an important step in the process of obtaining medium-term forecasts of GDP growth.

59. Because potential output can never be observed, estimates of it must be inferred from other data and information about the economy. Exactly how this is done can lead to different outcomes. How it *should* be done is not a resolved issue, even in principle; the choice is likely to depend on the structure of the economy as well as on views about what the concept of potential output actually tries to measure.

60. Thus it is perhaps not surprising that a number of different methods have been proposed in the literature and used at the IMF. As a prelude to the discussion, in Section IV below, of the methods used by IMF economists, this section introduces the different classes of methods available. It underlines the need for caution in the use of the Hodrick-Prescott

[28] At time t, once an estimate of the level of potential GDP is obtained for the desired horizon—say k periods into the future—and the actual level of GDP is known up to time t, the analyst can assess the cyclical position of the economy by looking at the difference between the actual and potential GDP (i.e., the output gap) at time t. Typically, it is assumed that at some point in the future (not necessarily between t and $t+k$) the effects of the cyclical factors (which ultimately drive the actual GDP away from the long-run trend given by the potential GDP) will vanish and the output gap will close. This provides a way to infer the speed of convergence of actual towards potential GDP that can be used to forecast the GDP growth k periods ahead.

Filter, which is popular among IMF desk economists. Annex 1 provides a fuller description of the various methods and discusses their advantages and disadvantages.[29]

61. Methods of estimating potential output fall into three broad classes, described in turn in what follows. At one end of the spectrum are purely statistical univariate approaches, which only use information contained in the GDP series itself to generate estimates of potential output. These are "mechanical" methods that have no economic content. At the other extreme are structural methods that rely on a fairly detailed model representation of the economy and typically use data from several variables that are explicitly included in the underlying economic model and in the estimation process. In between these two polar cases are bivariate and multivariate approaches—typically filters derived from purely statistical univariate methods that are modified to include some restrictions from economic theory, accounting identities, mechanical relationships between variables, and at least two observable variables in the estimation process.

A. Statistical Univariate Methods

62. Univariate methods use only the time-series of real GDP to produce estimates of potential output and the output gap, typically as a residual. This class of methods includes simple extrapolations from linear trends, moving averages, and other polynomials used to represent past observations. It also includes methods such as the Beveridge-Nelson decomposition, the very popular Hodrick-Prescott (HP) Filter, band-pass filters, and univariate versions of state-space and regime switching models.

63. The main shortcoming of univariate methods is that they are purely statistical, mechanical filters. Because they do not use economic theory in the identification of cycle and trend, they have difficulty in separating the effects of one-off events (the completion of an infrastructure project, for example) between trend and cycle and must be heavily complemented by the forecaster's overall knowledge of the country and by personal judgment. A second major problem of some of the methods in this class (e.g., the HP filter) is that they do not provide measures of the uncertainty around the estimates (Kuttner, 1994).

64. Although univariate methods admit modifications—bivariate or multivariate versions that make use economic theory to extract information about the cycle that is possibly contained in other variables than GDP itself (e.g., inflation, unemployment, capacity utilization, financial variables etc.)—their lack of economic content means, for example, that they do not take account of the relationship (known as the Phillips Curve) thought to exist between inflation and the slackness in the economy (i.e., the output gap) or unemployment.

[29] See also De Masi (1997), Kuttner (1994), Cerra and Saxena (2000), D'Auria and others (2010), Benes and others (2010), Morley and Piger (2012), Borio, Disyatat, and Juselius (2013), and Johnson (2013), Johansson and others (2013), among others.

These relationships are typically considered in macroeconomic models and are central to the definition of potential output as the maximum production without inflationary pressure.

65. An advantage of these methods, on the other hand, is that they only require data on the GDP, which makes them very easy to implement for a wide range of economies. The results are also straightforward to interpret. The HP filter, for instance, is a very popular method in general, and is widely used by IMF desk economists, according to interviews conducted by the evaluation team.

66. *Hodrick-Prescott (HP) Filter*. Because of its popularity among IMF economists, a brief discussion of two major issues associated with the HP filter is useful in relation to the sometimes large revisions often observed in past estimates of potential output, as discussed in Section II above. The first issue is the high sensitivity of the estimation of potential output to the arbitrary assumptions about the smoothness of the trend that are required by the HP procedure. Assuming a trend that is too smooth has implications for the volatility of the estimated output gap: it can generate excessive cyclicality in the output gap when there is little, or smooth out existing structural breaks in potential output. The second major drawback of using the HP filter to estimate potential output is its high sensitivity to the addition of new data points at the end of the sample period.

67. To illustrate, we consider the annual data on real GDP for the Unites States published by the Federal Reserve Bank of St. Louis. Figure 6 shows three series of estimated growth rates of potential output using information from 1980 to 2012. The blue and red lines refer to the estimation of potential GDP growth using a smoothing parameter, as required in HP filtering, that is arbitrarily set at $\lambda = 100$. The green line refers to an estimation based on $\lambda = 6.5$. Both these values are commonly discussed in the literature.[30] The difference between the blue and red lines is that the former refers to an estimation using only data available up to 2007, while the latter uses the full sample up to 2012. The differences in growth rates are striking, especially at the end of the sample period.

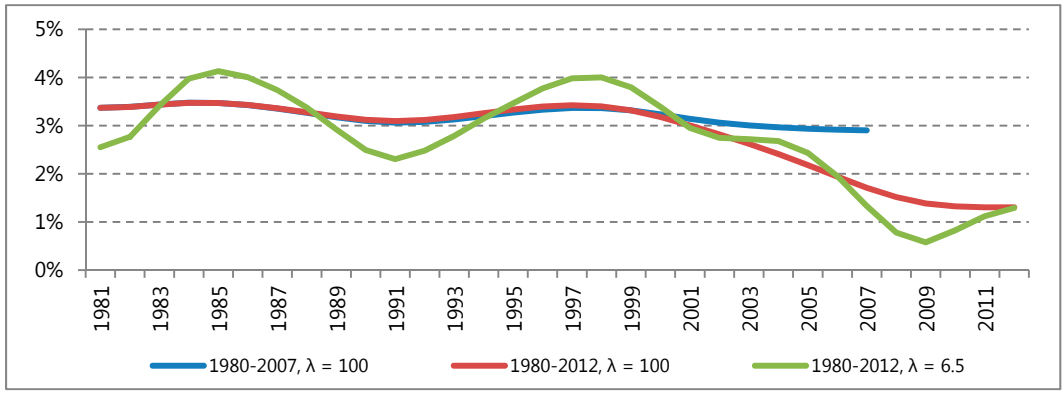
68. If the forecaster assumes a less smooth trend, as shown by the green line in Figure 6, the estimated growth rates of potential output start showing some cyclical behavior and, for some years, differ greatly from the cases with smoother trends.

69. More importantly, forecasters using this method in 2008 would have based their projections on a growth rate of potential GDP of about 3 percent in 2007. The same method, after including five extra observations, for 2008–12, would revise the estimates for 2007 and

[30] The value $\lambda = 400$, also very popular, produces very similar results to $\lambda = 100$. As λ gets very large (tends to infinity), the HP-trend converges to a linear trend. Hodrick and Prescott (1980, 1997) suggested a value of λ that is based on arbitrary prior views about what are the "normal" values of both the volatility of the cyclical component and the change in the growth rate of output within a quarter. See also Ravn and Uhlig (2002), Pedersen (2001), and Annex 1 below.

show a potential GDP growth of less than 2 percent. Note that this "end-of-sample problem" occurs precisely where accuracy matters most for forecasting and policy decisions. Moreover, the estimates of potential growth going back to 2001 would be revised after the introduction of new observations. This high sensitivity to new data may be a factor in similar revisions of past estimates that are common in IMF estimates of the potential GDP and output gap.

Figure 6. HP-based growth rate of U.S. potential GDP



Source: Author's calculations using data from the Federal Reserve Bank of St. Louis.

B. Multivariate Filters

70. Bivariate and multivariate filters offer a way to compromise between the greater theoretical coherence and data intensity of structural methods, discussed in the next subsection, and the simplicity and lack of economic content of univariate methods.

71. The first vintages of multivariate filters, such as that proposed by Laxton and Tetlow (1992), added economic theory to the otherwise mechanical or purely statistical methods by using the inflation-output gap nexus given by the Phillips Curve. The idea is to use the informative content of the inflation rate about the cyclical state of the economy (that cyclically high rates of inflation would be correlated with the cyclical component of output, the output gap) to better identify the potential GDP. Rather than analyzing the supply side directly, these methods use the joint behavior of output and inflation, informed by economic theory, to construct estimates of potential output and give the output gap more economic content, consistent with the conceptual definition of potential output based on stable inflation.

72. Not only inflation but other variables such as unemployment and capacity utilization can also be incorporated in multivariate approaches. See Cerra and Saxena (2000) for additional examples and Borio, Disyatat, and Juselius (2013) for models with financial variables.

73. At the IMF, Benes and others (2010) propose a model that, in addition to the Phillips Curve, also incorporates information about the cyclical state of the economy extracted from unemployment and capacity utilization rates. Using this information helps ensure consistency

between the estimated trend and cycle and these indicative variables of the business cycle. Their method, which is often used by Fund economists, and other methods in the same class of models, also help to reduce (though they do not eliminate) the common "revisions of the past" observed in estimates of potential GDP based on the HP filter.

C. Structural Methods

74. Structural approaches include the production function (PF) method, dynamic stochastic general equilibrium (DSGE) models, and structural vector-autoregressive models (SVARs). These methods try to construct measures of potential output that are consistent with a larger set of restrictions from underlying economic models, which can vary in detail and complexity, and typically use several variables during the estimation process.

75. Like the multivariate filters, structural methods also mitigate some of the problems associated with univariate statistical procedures concerning end-of-sample bias. However, by relying on much more detailed models of the economy, they often go beyond the multivariate filters' relatively crude representations of aggregate supply and demand forces. The increase in detail, especially in the case of DSGE models, provides a more solid economic basis for estimating potential GDP but it also makes these methods much harder to implement.

76. The main advantage of structural approaches is, therefore, the clear economic interpretation of the evolution of the estimated output gap combined with a trend that is consistent with that interpretation. However, model misspecification can be a major problem. As pointed out by Borio, Disyatat, and Juselius (2013), these methods' advantage in terms of theoretical coherence is only meaningful to the extent that the underlying models are good approximations of reality.

77. *The production function (PF) method*. Within this class the PF method deserves special attention, since it is the preferred method in other multi-country institutions such as the European Commission (EC) and the Organization for the Economic Cooperation and Development (OECD) but is not widely used by IMF economists.

78. The PF approach postulates a specific relationship—a production function—between the output and the inputs, or production factors (labor, physical and human capital, etc.). The growth of output reflects either the observed accumulation of production factors or the increase in their productivity (total factor productivity, TFP), which is unobserved and must be estimated (usually as a residual). Estimates of the trends in inputs, which are often obtained using univariate or multivariate filters, are plugged into the production function to generate the potential output.

79. The PF method can be applied to a single country, as in Johnson (2013), without considering either spillovers from other economies or information from the history of similar economies in terms of forces driving the secular trend in GDP; or it can be combined with

insights from economic growth theory and cross-country information to incorporate "catch up" factors that may drive per capita income to converge across similar, peer economies.

80. The European Commission (EC) uses a comprehensive version of the PF approach, described in D'Auria and others (2010), that combines bivariate and multivariate filters to bring economic theory to bear in the identification of the trend in inputs and TFP. In the EC, the use of the PF method ensures homogeneity of treatment to all member countries—which is required because potential output and the output gap are needed to compute the cyclically adjusted fiscal balances that are formally monitored by the Commission.

81. Johansson and others (2013) propose a version of the PF approach, currently used for long-term forecasts at the OECD. Their model uses the notion that (per capita) output converges, in the long run, to a balanced growth path determined by the interplay between global technological progress (which makes the "technology frontier" advance) and country-specific structural conditions that include demographics, institutions, and policies. The speed of convergence is typically governed by the distance from the technology frontier and, to some extent, by the rates of accumulation of production factors such as labor and human and physical capital.

82. By allowing a "catch-up" mechanism—whereby countries that are relatively similar, conditionally on the available domestic factors, converge to approximately the same growth rate of potential (per capita) output—and by also taking careful consideration of demographic trends, the OECD's version of the PF approach provides a consistent cross-country framework for the analysis of long-term growth and, therefore, seems very relevant to multi-country institutions such as the IMF.

83. The conditional convergence of income across similar countries seems to contain information about the medium- and long-term path of GDP that is not always used by economists who apply the PF approach to a single country.

84. This is especially noteworthy because the centralized, top-down coordination process observed in the IMF's short-term forecasts (see Genberg, Martinez, and Salemi, 2014)—does not seem to be as effective in the context of medium-term forecasts (as discussed in Section IV below). Batista and Zalduendo (2004) and IMF (2004) have considered the use of a similar approach at the IMF. Their results indicate that, especially in developing economies, the information about conditional convergence and other (institutional, policy-related) factors that help explain growth in a cross-country panel can improve the accuracy of medium-term forecasts of GDP growth.

85. For some types of economies, however, structural approaches that take into account economic relationships such as the Phillips Curve may not be very suitable for medium-term forecasting of GDP. These cases include those economies that do not have a well developed industrial sector, are mainly exporters of agricultural or natural resources, or have a long-lasting fixed exchange-rate peg that essentially ties domestic inflation to foreign prices.

Moreover, structural methods and multivariate filters may not be appropriate for economies where a large proportion of output is supply-determined and highly dependent on the evolution of prices that are determined outside the country (as for commodities, oil).

IV. The IMF's Medium-Term Forecasting Process and Methods

86. This section discusses aspects of the IMF forecast process that are specific to medium-term forecasts and have not been covered in detail in the general description available in Genberg, Martinez, and Salemi (2014). We draw on evidence from the survey and interviews with desk economists regarding their use of specific methods to estimate potential GDP. The focus is on the production of point forecasts for the medium term.

A. Process

87. The process used by the IMF to conduct forecasts encompasses both short- and medium-term forecasts. Forecasts for all horizons are integrated in the Fund's spreadsheet-based macro framework. This framework[31] is a set of relationships among economic variables built into spreadsheets that jointly describe a country's flow of funds. Country desk economists employ these spreadsheets to organize information for their routine analysis of the economy and to support forecasts. The relationships primarily consist of macroeconomic accounting identities but may also include a small number of behavioral equations and arbitrage conditions. They ensure that the point forecasts are consistent over the forecast horizon (i.e., up to five years ahead) and with regard to the economy's flow of funds between the monetary, fiscal, real, and external sectors. Interviews with staff showed that the use of the macro framework is country-specific and varies greatly in detail and sophistication, ranging from the use of "satellite" models—to forecast certain parts of the spreadsheet—to simply entering numbers based on judgment.

88. The main differences of medium- relative to short-term forecasting at the IMF are the stronger emphasis on the supply side, the important role of estimates of potential output, and the lesser degree of coordination across country desks. Just as for short-term forecasts, desk economists are responsible for making the forecasts (under supervision) and there are no formal guidelines or top-down directives specifically on how they should do so. Medium-term forecasts are subject to the same constraints and mechanical consistency checks[32] imposed by the *WEO* process on short-term forecasts. But in the interest of flexibility, there are no considerations or incentives for methodological consistency, even among countries that broadly share similar characteristics.

---

[31] Formerly termed the "financial programming framework."

[32] These checks aim at ensuring that accounting identities and standard theoretical presumptions are not violated, detecting possible reporting errors (misplaced decimal places, unintended changes in units of measurement, etc.), and flagging any large changes in forecasts relative to previous forecasts.

89. Evidence from the interviews with IMF staff, mostly desk economists, indicates that efforts to construct an analytically unified view of the global economy over the medium term and to ensure regional and global consistency are much less noticeable, if they exist, than for short-term forecasts. The combination of top-down and bottom-up aspects of IMF forecasts is much more skewed towards the latter in the case of medium-term forecasts and, differently from short-term forecasts, top-down elements are largely absent.[33] In particular, there is no interdepartmental committee in charge of maintaining analytical consistency among medium-term forecasts at the country, regional, and global levels, as the Interdepartmental Forecast Committee does in the case of short-term forecasts.

90. The lesser degree of coordination across different country desks and area departments in the context of medium-term forecasts may explain why some signs of informational inefficiency are more common in these forecasts than in their short-term counterparts, as discussed in Section V below.

B. Methods Used by IMF Desk Economists

91. The survey and post-survey follow-up interviews with IMF staff, as well as analysis of Fund documents, indicate that all the methods for estimating potential output described earlier in this section are or have been used, to some extent, by IMF economists to conduct medium-term forecasts of GDP growth. All the methods have shortcomings and the appropriate choice among them depends on data availability and on the structure of the economy in question—including how it changes over time, through successive development stages or structural reform.

92. Survey evidence shows that in the Fund's medium-term forecasting the use of any particular individual forecasting method is much less universal than the use of judgment—understood as a set of information and knowledge, not necessarily quantitative in nature, that desk economists and mission chiefs accumulate about the countries on which they work. Table 2 summarizes the staff responses to a question about the methods—broadly defined—used to conduct medium-term forecasts at the IMF. Desk economists were asked to indicate which methods they use by selecting among structural models, statistical methods, and judgment.[34] They were allowed to choose more than one option.

93. Judgment is the most frequent method, and is used by almost four out of five desk economists. Structural models and statistical methods are used much less frequently (34 percent and 44 percent, respectively). This finding is consistent with Genberg, Martinez,

[33] Genberg, Martinez, and Salemi (2014) describe the coordination efforts in the context of short-term forecasts.

[34] The term "structural" has a strict and a broader interpretation. The strict definition restricts structural models to models, such as the DSGE models, that have micro foundations derived from first principles (assumptions about rational behavior and technology). The broad definition includes other types of economic models. The evaluation team did not elaborate on which definition was been considered and left the staff free to decide.

and Salemi (2014) and is robust: the widespread use of judgment holds in different cuts of the data across IMF area departments and levels of economic development.

Table 2. Methods used by IMF desk economists in medium-term forecasts
(In percent)

	Structural	Statistical	Judgment	Other
Full sample	34.1	44.4	78.6	9.5
APD	47.1	47.1	76.5	11.8
AFR	14.8	33.3	74.1	7.4
EUR	56.3	50.0	78.1	9.4
MCD	22.7	36.4	90.9	9.1
WHD	28.6	53.6	75.0	10.7
ADV	75.0	45.0	70.0	10.0
EME	32.8	46.9	75.0	7.8
LIC	16.7	40.5	88.1	11.9
OIL	22.2	50.0	77.8	5.6
Training				
Yes	35.4	50.0	79.2	8.3
No	33.3	41.0	78.2	10.3
Experience				
High	33.3	44.4	83.3	11.1
Low	34.3	44.4	77.8	9.3

Source: Author's calculations using IEO Forecast evaluation survey data.

Note: Desk economists are classified as having "high" experience if they have been working on their current desk for three years or more and have worked on at least three country desks, abstracting from the experience acquired outside the IMF.

94. Not surprisingly, the use of judgment is more frequent among desk economists working on countries at lower levels of development, although as many as 70 percent of the economists working on advanced economies use judgment, too. The opposite is true for structural models: fewer than 15 percent and 17 percent of the staff making forecasts for African and low-income countries, respectively, use structural models, compared with 75 percent of those forecasting advanced economies.

95. Table 2 also suggests that on-the-job training in forecasting makes little difference to the choice of methods: economists who attended IMF-provided courses or seminars specifically related to forecasting are equally likely to use judgment and structural models relative to staff who did not attend these training events, although training seems to favor the use of statistical methods. Nor is experience a strong factor in the choice of methods, although more experienced economists slightly prefer to use judgment.

Use of judgment

96. The interviews show that judgment is applied in different ways. Some desk economists apply relatively sophisticated models to forecast the short term, and then use judgment to form a "reasonable" guess about the sustainable growth rate in the very long run

to interpolate the short-term with three-, four-, and five-year-ahead forecasts. Others start with the long run, often using one of the methods to estimate potential GDP as a guide, and then go back to construct short-term forecasts that are consistent with the forecasts for longer horizons. In all cases, the macro framework ensures that the forecasts are consistent across different horizons and variables.

97. As suggested several times during interviews, the heavy use of judgment by desk economists may reflect unavoidable data availability issues and frequent structural changes in member economies that prevent the use of more sophisticated methods when working with less developed economies. In addition, the large proportion of IMF member countries facing structural change may limit the use of quantitative models that are more suitable for advanced and mature economies. It is also likely that IMF economists, even when data are available, appropriately apply judgment rather than purely relying on the quantitative method of choice.[35]

Use of quantitative methods

98. Data availability is a major impediment to the use of structural methods. Evidence from survey and interviews with staff indicates that the production function (PF) method is currently little used at the IMF, mainly because its use calls for data that are frequently unavailable for many member countries.[36]

99. The use of the PF method, like most quantitative methods, is further complicated when applied to economies experiencing large inflows or outflows of capital and labor, facing structural change, or in demographic transition, as may be the case for a large share of the IMF membership. These factors make the identification of trends in both inputs and total factor productivity much more difficult; they mask the effects of one-off structural breaks, and increase the likelihood that desk economists must rely mainly on their judgment to estimate potential output and forecast GDP over longer horizons. Insights from empirical studies on the sources of economic growth may help in these cases.

100. Post-survey interviewees also explicitly mentioned other methods for estimating potential output and for guiding medium-term forecasts of GDP growth: (a) the averaging of estimates of potential output produced by several methods—the production function method, regime switching models, and state-space models, as in Johnson (2013); (b) the multivariate filter developed at the IMF Research Department by Benes and others (2010); and (c) the use of simple averages and trends based on past data.

---

[35] This suggests that if a readily available forecasting "toolbox" is made available (as discussed in Subsection C below), it will not preclude the appropriate use of judgment.

[36] De Masi (1997) suggests that staff may have used the PF method—combined with insights from cross-country empirical growth models—more often in the past.

101. However, perhaps because it is easy to apply, the HP filter, alone or in combination with other methods and/or judgment, was the quantitative method most frequently mentioned by interviewees (25 percent). Some staff expressed concern that the HP method may be used too mechanically and without due consideration for the uncertainty around its trend-cycle decomposition. Widespread mechanical use of the HP filter could help explain the frequent revisions of past estimates of potential GDP observed in the *WEO*, as discussed in Section II above.

102. To assess, albeit imperfectly, the role of mechanical methods in the IMF five-year-ahead forecasts of GDP growth, Table 3 displays the correlations between IMF forecasts and forecasts based on a series of "naïve" forecast methods.[37] A positive correlation is consistent with the assumption that these methods are used by IMF economists.

Table 3. Correlation between Spring *WEO* forecasts and naïve forecasts

	HP	Linear trend	MA(5)	RW
		Spring		
5 Years				
Correlation	0.03	0.01	0.10	0.03
p-value	0.07	0.69	0.00	0.05
Obs	4017	4017	3832	4017
4 Years				
Correlation	0.06	0.02	0.16	0.06
p-value	0.00	0.17	0.00	0.00
Obs	4018	4018	3833	4018
3 Years				
Correlation	0.09	0.03	0.17	0.07
p-value	0.00	0.09	0.00	0.00
Obs	4018	4018	3833	4018
		Fall		
5 Years				
Correlation	0.04	0.01	0.12	0.04
p-value	0.02	0.53	0.00	0.01
Obs	4024	4024	3835	4018
4 Years				
Correlation	0.06	0.01	0.15	0.05
p-value	0.00	0.68	0.00	0.00
Obs	4024	4024	3835	4018
3 Years				
Correlation	0.07	0.02	0.16	0.05
p-value	0.00	0.15	0.00	0.00
Obs	4024	4024	3835	4018

Source: Author's calculations using *WEO*.

[37] The naïve methods are the HP filter, linear trend, 5th-order moving average, and a random-walk model, mechanically constructed so as to replicate the conditions faced by a forecaster in real time. More specifically, a forecast for year $t+k$ made at year t only uses information available up to that time; in this case, only data up to year t-1 available at the *WEO* vintage released at year t. For next year, a new forecast is obtained using only information up to year t, available from a subsequent vintage of the *WEO*, and so on. For each year t, the k-year-ahead forecasts are the forecasts of the GDP growth rate k years from year t, rather than the annualized rate of GDP growth over the next k years.

103. The correlations in Table 3 are always positive and strongly statistically significant, except in the case of the linear trend method, suggesting that naïve methods may have been used, at least partially, to construct medium-term forecasts of GDP growth at the IMF.[38, 39]

104. Not surprisingly, though, the correlations are not very high. This could suggest the use of more sophisticated methods than the naïve approaches considered but, in light of the results in Table 2 above, it is likely to simply reflect the widespread use of judgment by IMF economists. Section V below shows that IMF medium-term GDP growth forecasts, perhaps because of the value added by staff's judgment, are clearly more accurate than the forecasts yielded by these mechanical methods.

105. Table 3 also shows that the correlations are higher for forecasts with shorter horizons, suggesting that the value added by the staff's approach, relative to mechanical methods, may increase as the forecasting horizon lengthens. For instance, forecasts based on a 5th-order moving average process at the three- and five-year horizons and published in the Spring vintage of the *WEO* show correlations of 0.17 and 0.10, respectively.

106. All in all, the evidence in Table 3 suggests that mechanical methods may be used in medium-term forecasting of GDP growth at the IMF, but are far from providing a full explanation of these forecasts. Combined with the staff survey results (Table 2), the evidence is consistent with the notion that IMF desk economists complement their chosen forecasting methods with a great amount of judgment, as it has been largely confirmed during the interviews with staff.

C. Discussion: Choice of Forecasting Methods

107. The variation in the choice of medium-term forecasting methods partly reflects the heterogeneity among IMF member countries in data availability, structural change, and development stage, but it may also be explained by the lack of a more centralized, top-down approach to medium-term forecasting, as used in IMF short-term forecasts and by other multilateral institutions with less diverse membership than the IMF, such as the European

[38] The statistical significance can be assessed based on the *p*-values reported in Table 3, which are a measure of how likely the data are to have been generated by chance assuming that the (null) hypothesis of a zero correlation is true. A *p*-value of 0.07 for example, indicates that the difference between the observed correlation and zero is large enough so that the maximal significance level—i.e., the probability of rejection of the hypothesis of zero correlation when it is fact true—consistent with the rejection of the hypothesis is 7 percent. If the analyst is comfortable with, say, at most 5 percent, the hypothesis of a zero correlation cannot be rejected, but if the analyst's comfort level is higher than 7 percent, then that hypothesis must be rejected. Rejecting the hypothesis of a zero correlation means that the observed value is deemed statistically different from zero and it is, therefore, considered statistically significant.

[39] Note the *p*-values usually lower than 0.05, and always lower than 0.1, except for the case of a deterministic linear trend. These are standard levels of significance usually considered in the literature.

Commission and the OECD.[40] Except for the mechanical checks used in *WEO* forecast rounds (see footnote 32), little attention is paid to economic and analytical consistency across countries in medium-term forecasts relative to short-term forecasts.

108. Leaving desk economists free to choose their forecasting methods is not a problem, provided that they are well informed about the methods available for the task and the best way to use them in different countries/situations and—especially—if they have the right incentives to seek information about these methods in order to improve their forecasts. Under these ideal situations, the flexibility allows the desk economists to perfectly adapt their forecasting approach to the (definitely heterogeneous) circumstances of their country of assignment.

109. Interviews with staff indicate, however, that these conditions are not always in place and that having some flexible guidance on methods could be desirable from their perspective. A more unified approach to medium-term forecasting could not only help desk economists who lack the time or incentives to look for a better method to conduct forecasts but could also induce the use of similar classes of methods in forecasts for similar economies (those facing the same technical and data restrictions, experiencing similar development phases, or having the same status regarding IMF program participation) and thus increase the analytical consistency of forecasts. Such consistency could be important in products that rely on relative forecasts of GDP growth, such as the Pilot External Balance Assessment. It may also produce efficiency gains by fostering IMF economists' use of forecasts for other economies when forecasting for their country of assignment.[41]

110. Another potential argument for a more unified approach to medium-term forecasts is the heavy reliance of IMF medium-term forecasts on staff judgment in a situation of high turnover rates among desk economists.[42] Such an approach could mitigate the loss of information due to staff turnover by guiding new desk economists on the choice of methods for their specific economy and/or guiding the development of their judgment.

---

[40] At the EC, a central unit coordinates the efforts of the teams involved in producing medium-term forecasts for individual economies, with a view to assure not only accounting, technical, and statistical consistency, but also consistency in the analytical and economic sense. At the OECD, which also favors a more homogeneous approach, the top-down aspect of medium-term forecasts is also stronger than at the IMF.

[41] See the discussion about economic interdependencies in Section V, subsection C, below.

[42] Past evaluations by the IEO have repeatedly pointed to the deleterious effects of high turnover and the ad hoc nature of staff succession arrangements on the accumulation of country-specific expertise. See IEO (2009, 2011, and 2013). In 2013 the IMF Strategy & Policy Review Department issued an internal checklist/guidance note for country assignment handover within the department, to ameliorate the handover process (www-intranet.imf.org/departments/SPR/OGR/Pages/default.aspx). It is too early to tell whether this will have the hoped-for effects, and whether the same approach will be implemented also in area departments.

111. In post-survey interviews, IMF staff reacted favorably to the idea of having a somewhat more centralized process and unified approach for medium-term forecasts, as long as sufficient scope to accommodate cross-country heterogeneity at the IMF is maintained. This approach could be centralized under the top-down guidance of an oversight committee (possibly interdepartmental).

112. Also, several respondents mentioned that a ready-to-use "toolbox" containing some of the available forecasting methods—accompanied by guidelines and suggestions of how and in which circumstances each method should be used—could be very useful. Staff were of the view that, if guidelines about methods are provided, they should be flexible enough to accommodate cross-country heterogeneity. Staff also expressed concern that toolboxes can be misused, making forecasting too mechanical. On the other hand, such a toolbox could dramatically reduce the cost to economists of testing different models that may potentially lead to better forecasts.

113. A surprising result from both the survey and interviews of staff is that checking the performance of the method of choice, by comparing forecasts and out-turns, is *not* common among IMF desk economists.[43] Without an assessment of past forecast errors and incentives to search for better methods to improve forecasts, some inertia in changing forecast procedures should be expected. Such inertia was often mentioned by desk economists during interviews.

114. Nevertheless, the methods used and the general approach to medium-term forecasts at the IMF can only be fully assessed after due consideration of their quality based on a formal comparison with out-turns, which is the object of the Section V.

D. Summary of Findings on Process and Methods

115. The findings in the last two sections may be summarized as follows:

- There are several methods available to estimate potential GDP, which guide medium-term forecasts of GDP growth. All methods have shortcomings and no method is appropriate for all economies.

- Data availability and differences in the structure of the economy being forecasted should and, to some extent, do guide the choice of methods by IMF desk economists.

- Some of the filtering techniques used to estimate potential GDP, including the HP filter (popular among IMF economists), are very sensitive to the addition of new data to the sample. This undesirable property can not only affect current estimates, but

[43] See Genberg and Martinez (2014, a; b).

may also explain the often large revisions of past estimates of potential GDP growth found across different vintages of the *WEO*.

- There is little institutional guidance on how desk economists should produce medium-term forecasts. Medium-term forecasts are subject to many of the same mechanical checks imposed on short-term forecasts in the context of the publication of the *WEO*. But they are not subject to the same degree of coordination to ensure global and regional consistency in terms of their analytical and economic contents.

- Forecasts produced by purely mechanical methods, with no economic content, are positively correlated with *WEO* forecasts. The correlations are clearly statistically significant but not very high, suggesting that these methods do not fully account for the approach used by desk economists to medium-term forecasting, which relies much more heavily on judgment than in any particular quantitative method—structural or purely statistical.

- Overall, the methods used in medium-term forecasts of GDP growth for individual economies seem broadly appropriate, but sufficient coordination for better analytical consistency at the regional and global levels may be lacking, and it is not clear whether the right incentives are in place for the systematic use of the "best" method for each specific economy.

V. Performance of IMF Medium-Term Forecasts

116. Like most economic forecasters, the IMF failed to foresee the large drop in real GDP growth in 2008–09 that took place in many of its member countries. Large forecast errors in high-profile cases that ended up in crises and required the use of IMF resources—the 1997 Asian Crisis, Argentina in the early 2000s, and Greece, more recently—have not been uncommon either. Some observers have used these episodes and other large forecast errors in individual countries to question the accuracy and, in some cases, the integrity of IMF forecasts.

117. This section assesses the performance of IMF medium-term forecasts of real GDP growth, based on three separate metrics: informational efficiency, both reflected in biased forecasts (Subsection B) and of more general nature (Subsection C), as well as accuracy (Subsection D).[44] Subsection A introduces the data and scope of the analysis, and Subsection E summarizes the findings. Opinions about the forecasts, as measured by survey and interview data from users and staff, are analyzed in Section VI.

[44] The main report (IEO, 2014) also mentions a fourth metric, the ability to learn from past forecast performance, to assess the quality of forecasts in general (including medium-term forecasts).

A. Quantitative Analysis of Forecasts: Data and Scope

118. The analysis here broadly follows the methodology that Timmermann (2006) and Musso and Phillips (2002) used to analyze short-term forecasts, but it focuses on medium-term—i.e., three-, four-, and five-year-ahead—point forecasts of the annual growth rate of real GDP.

119. All IMF member countries for which there are more than six observations are included in the sample. In some cases, the sample is trimmed to account for outliers and other potential problems with the data. Annexes 2 and 3 provide a detailed account of the methodology and the findings.

120. For a forecast made at year t, for year $t+k$, the forecast error is the difference between the actual realization (out-turn) of the growth rate at $t+k$ and the forecast. Negative values are therefore associated with optimistic forecasts. We consider both vintages of the *WEO*—Spring and Fall—separately and, to accommodate revisions in the data, we use values of out-turns that are measured with delays of one and two years.

121. The assessment covers the period 1990–2012 using data from the *WEO* database.[45] Recent years are not part of the sample because it is still too early to calculate their forecast errors. Therefore, differently from recent studies of short-term forecasts, predictions that were made during the recent financial crisis are mostly excluded from the analysis.[46]

122. Under the assumption that the forecaster knows the structure of the economy, typical tests of forecast informational efficiency seek to assess whether forecasts are unbiased in the statistical sense (i.e., whether the average forecast error is zero), not serially correlated (i.e., past errors are not correlated with future errors), and whether errors cannot be predicted by making use of any information available to the forecaster at the time of the forecast.

123. If these conditions are not satisfied, forecasts can be improved by simply adding the observed bias to the forecast and taking into account the information content of serially correlated past errors or of whatever other variable may help predict errors. If forecasts can be improved along these lines, then they are considered inefficient because some available information is being "wasted" (i.e. not fully used to improve the forecasts).

124. To assess whether IMF medium-term forecasts in the *WEO* are efficient we follow two approaches: (i) a country-by-country analysis of forecast errors, based on both

[45] The first observations available on five-, four-, and three-year-ahead forecast errors (based on forecasts made in 1990) are computed for 1995, 1994, and 1993, respectively. The last observations are recorded for 2011, for forecasts made five, four, or three years earlier.

[46] With the exception of three-year-ahead forecasts that were made in 2008 for GDP growth in 2011, and whose out-turn was measured in 2012.

descriptive and regression-based statistics, and (ii) an aggregate analysis that relies on pooling forecast errors across countries and over time.[47]

125. Some caveats should be noted before considering the results. First, as argued by Faust (2013), statistical tests of unbiasedness, accuracy, and overall efficiency may be a poor assessment of the quality of forecasts for economies in which there are relatively frequent structural changes.[48]

126. Second, our tests of forecasts refer to point forecasts. Looking at the entire distribution of possible outcomes, perhaps with the help of risk scenarios, may be more appropriate in some cases, especially for medium-term forecasts. Results from the survey conducted by the evaluation team suggest that risk analyses and scenarios are sometimes more valued by country authorities than point forecasts.

127. Third, several of the results discussed below are not robust to generalizations; they vary with the measure of the bias (mean or median) and are affected by considerations related to sample size, forecast horizon, and *WEO* vintage (Spring or Fall), especially in light of the recent financial crisis.[49] Changes in assumptions along those dimensions affect the sample and may greatly change the results. In addition, many of the results may be subject to small-sample distortions, despite our efforts to account or correct for them.

B. Are Forecasts Biased?

128. The main findings regarding the degree of bias in forecasts, according to the country-by-country and the aggregate approaches, are summarized in Tables 4 and 5, respectively. Recall that negative values are indicative of optimistic bias—i.e., overprediction of GDP growth—while positive values indicate a pessimistic bias.

129. Two measures of the bias are considered in the descriptive analysis—the mean forecast error and the median forecast error. The median-based measure of bias provides a

[47] Both the regression-based country-by-country analysis and the aggregate approach are based on the regression of forecast errors—in each individual country, in one case, and from the pooling of data across countries, in the latter case—on a constant. In the country-by-country analysis, a second specification includes lagged forecast errors in the regression to detect serial correlation.

[48] Efficiency tests, similar to those used in this section, rely on the assumption that the forecaster knows the structure of the economy being forecasted. If there are "learning opportunities" for the forecaster (for example, as structural reforms are laid out and their effects are only gradually felt in the economy), a "failure" in these tests may not be caused by factors available to the forecaster that are not considered in the forecasts, but may simply be a result of the time needed for the forecaster to catch up with the changing structure of the economy and for this learning to be gradually reflected in the sample used in the tests.

[49] See Genberg and Martinez (2014b) for similar remarks regarding short-term forecasts.

more robust measure of the center of the errors' distribution, less sensitive to outliers.[50] Only the mean forecast error is obtained in regression-based tests and in the aggregate analysis.

130. We find that IMF medium-term forecasts of GDP growth tend to exceed out-turns, on average. The cross-country average of the mean forecast error ranges from -0.44 percentage points to -0.76 percentage points; using the median, biases vary between -0.14 percentage points and -0.41 percentage points.

131. Between two-thirds and three-fourths of countries are associated with negative values of either the mean or the median forecast error—a much higher rate than the expected 50 percent threshold of an unbiased sample (Table 4). But in only one-fifth to one-third of countries is there statistically significant evidence of bias, mostly (between 18 percent and 31 percent of countries) on the optimistic side.

Table 4. Country-by-country analysis: Biases in medium-term forecasts of GDP growth
(Range of results)

	5-year-ahead	4-year-ahead	3-year-ahead
	descriptive statistics		
Average mean bias (percentage points)	Between -0.44 and -0.56	Between -0.53 and -0.67	Between -0.64 and -0.76
Optimistic – total (Percent of countries)	72–73	73–76	74–78
Optimistic – statistically significant (Percent of countries)	18–25	21–26	23–31
Average median bias (percentage points)	Between -0.14 and -0.24	Between -0.18 and -0.34	Between -0.24 and -0.41
Optimistic bias – total (Percent of countries)	64–68	64–72	62–71
Optimistic – statistically significant (Percent of countries)	21–27	21–28	21–30
	regression-based statistics		
Average bias (percentage points)	Between -0.44 and -0.56	Between -0.53 and -0.67	Between -0.64 and -0.76
Statistically significant bias (Percent of countries)	24–33	22–28	19–28

Source: Tables A2.1 and A2.2, in Annex 2.

[50] The effect of outliers in the assessment of forecast biases can be seen in Table 4, which shows that the average cross-country bias as measured by the median is much smaller (sometimes less than half) than that measured by the mean, while the percentage of countries showing bias is about the same. This reflects the presence of outliers in the right extreme of the distribution of forecast errors, that increase the (absolute) value of the mean but not the median forecast error. Outliers, combined with differences in the sample period, partially explain why the cross-country average forecast error seems to increase for shorter forecast horizons. When controlling for the same sample, this difference is slightly reduced, but it virtually disappears when outliers and asymmetry in the distribution of errors are removed (by truncation). See Table A2.2, in Annex 2.

132. The absence of bias in the forecasts for the vast majority of countries does not support the notion of a built-in organizational bias, but the results point to potential problems in a non-trivial share of member countries, which may deserve attention from IMF Management.

133. The country-by-country results reported in Table 4 are obtained with fewer than 20 observations—and sometimes as few as 6. In these situations, small-sample distortions may apply.[51] When these distortions are taken into account, the frequency of statistically significant biases is further reduced.[52]

134. An alternative way of getting around the small-sample problem is to use pooled country data in the regression-based analysis. The trade-off is the loss of results for individual countries, keeping the focus only on aggregate results. Table 5 shows the results of this aggregate analysis.

Table 5. Aggregate analysis: biases in medium-term forecasts of GDP growth

	5-year-ahead	4-year-ahead	3-year-ahead
Average bias - Baseline	-0.3(*)	-0.4(*)	-0.5 (*)
Excluding " big" recessions	-0.0	Between -0.1 and -0.2 (*)	Between -0.2 and -0.3 (*)

Source: Table A2.3 in Annex 2. Results obtained using pooled data across countries. Numbers are rounded to one decimal place. (*) Indicates statistical significance at the 5 % level.

135. The overoptimism in medium-term forecasts of GDP growth is confirmed: the mean forecast error is negative and strongly statistically significant in all horizons. The average bias ranges from about -0.3 percentage points to about -0.5 percentage points. The tendency to larger optimistic bias in forecasts with shorter horizons, previously discussed, resurfaces.

The role of recessions

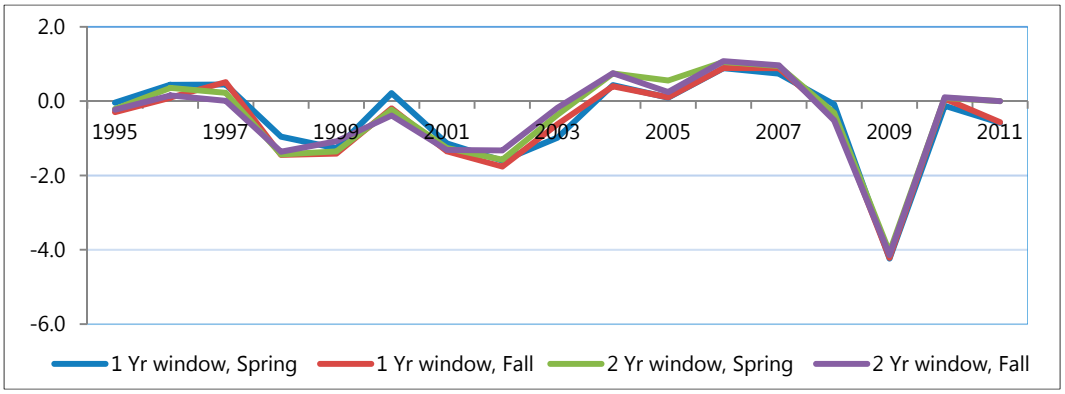
136. The measurement of biases may also be affected by the sample period. Figure 7 displays the year-by-year cross-country (simple) average of the mean forecast errors in five-year-ahead forecasts. Table 6 shows the same information also for other forecast horizons.

137. In our sample period, the frequency of optimistic cross-country average bias (i.e., negative values), is only slightly higher than that of pessimistic bias, which is heavily concentrated in the years 2004–07. However, in absolute terms, optimistic biases are larger than pessimistic biases, explaining the negative values when averages are computed over time (Table 6).

[51] For instance, small samples increase the chances that existing biases may not be picked up by the tests unless they are sizeable. Actual problems that are not extreme may go undetected. On the other hand, for a given average deviation of actual outcomes from the forecast, the likelihood of finding a statistically significant bias when there is actually none increases when there is less variability in the sample, especially if small.

[52] A bootstrap procedure is used. See Annex 2 for further details of the results.

Figure 7. Cross-country simple average of 5-year-ahead forecast errors
(percentage points)



Source: Author's calculations.

Table 6. Cross-country average of forecast errors in IMF medium-term forecasts of GDP growth
(Percentage points)

	5-Year Ahead				4-Year Ahead				3-Year Ahead			
	1 Yr window		2 Yr window		1 Yr window		2 Yr window		1 Yr window		2 Yr window	
Year	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall
1995	-0.04	-0.30	-0.21	-0.23	0.10	-0.26	-0.07	-0.19	-0.01	-0.49	-0.17	-0.42
1996	0.44	0.09	0.36	0.15	0.28	0.10	0.19	0.16	-0.16	-0.43	-0.23	-0.38
1997	0.45	0.51	0.22	0.01	0.02	0.04	-0.14	-0.37	-0.02	0.03	-0.21	-0.36
1998	-0.96	-1.45	-1.43	-1.36	-1.03	-1.46	-1.47	-1.35	-1.13	-1.67	-1.51	-1.59
1999	-1.27	-1.41	-1.35	-1.08	-1.38	-1.55	-1.46	-1.21	-1.56	-1.63	-1.63	-1.30
2000	0.21	-0.19	-0.23	-0.39	0.03	-0.32	-0.42	-0.52	-0.13	-0.47	-0.58	-0.69
2001	-1.14	-1.35	-1.26	-1.32	-1.21	-1.27	-1.35	-1.23	-1.42	-1.45	-1.52	-1.41
2002	-1.60	-1.76	-1.58	-1.33	-1.80	-1.89	-1.80	-1.45	-1.74	-1.70	-1.72	-1.27
2003	-0.98	-0.62	-0.37	-0.18	-1.21	-0.63	-0.61	-0.19	-0.96	-0.77	-0.35	-0.33
2004	0.43	0.40	0.74	0.75	0.16	0.63	0.47	0.99	0.41	0.51	0.72	0.86
2005	0.09	0.10	0.55	0.24	-0.23	-0.08	0.23	0.06	0.10	0.45	0.58	0.59
2006	0.88	0.89	1.05	1.08	0.76	0.84	0.92	1.04	0.76	1.29	0.93	1.48
2007	0.74	0.88	0.95	0.97	0.70	1.05	0.91	1.13	1.18	1.31	1.32	1.32
2008	-0.08	-0.41	-0.32	-0.52	0.00	-0.22	-0.23	-0.33	-0.24	-0.26	-0.47	-0.37
2009	-4.24	-4.22	-4.07	-4.14	-4.33	-4.21	-4.16	-4.13	-4.51	-4.54	-4.34	-4.46
2010	-0.13	0.06	0.09	0.10	-0.35	-0.29	-0.14	-0.25	-0.80	-0.66	-0.58	-0.62
2011	-0.59	-0.57			-0.90	-1.16			-1.34	-1.30		
Average	-0.46	-0.55	-0.43	-0.45	-0.61	-0.63	-0.57	-0.49	-0.68	-0.69	-0.61	-0.56

Source: Author's calculations using *WEO*.

138. It is notable that the start of spells of (large) optimistic bias seems to follow three important crises or recessions with global implications—the 1997 Asian Crisis, the 2001 recession associated with the terrorist attacks and the burst of the "dot-com" bubble in the United States, and the 2009 financial crisis. The large optimistic error of forecasts for 2009 is noteworthy. Similar patterns are observed in the four- and three-year-ahead forecasts (Table 6), suggesting that recessions may be a particularly important factor in the overall bias results, a notion confirmed by Genberg and Martinez (2014b) for short-term forecasts.

139. Excluding these large recessions from the sample eliminates the statistical significance of the estimated bias in five-year-ahead forecasts and greatly reduces the size of biases (by more than half) in both four- and three-year-ahead forecasts (Table 5 above). This result, too, is at odds with the notion of an organizational, systemic optimistic bias. It is also consistent with findings by Loungani (2001), who shows that biases in forecasts made by the private sector tend to disappear when periods of negative growth are removed from the sample.[53]

140. However, this result also indicates that an entrenched inability to predict crises or recessions is an important factor behind the observed bias in GDP growth forecasts. An inability to accurately predict recessions is not peculiar to IMF forecasters,[54] but still raises the question of whether the Fund should pay more attention to predicting recessions.[55]

141. The question of whether the IMF gives appropriate incentives to desk economists to improve forecasts came up during interviews with staff, mostly desk economists. Interviewees frequently pointed out that (i) good forecast performance is not typically reflected in their annual performance review (APR) and that (ii) the costs of deviating too much from available forecasts—whether their own past forecasts or those of peers—can be very high.

142. Interviewees reported that disagreeing with colleagues and mission chiefs on forecasts often requires great effort, which should not be necessarily problematic, but may introduce asymmetry in the IMF forecasters' behavior if there is no expected pay-off when they are proved right. If this thinking reflects the organizational culture regarding forecasts, it is easy to imagine a situation where desk economists are wary of reporting a forecast of a large drop in GDP following a period of stability, even if they have good reason to believe this will happen. Unless the evidence of a future large recession is overwhelming and/or the notion is widely shared by colleagues and pears, desk economists will underreport these cases and, if indeed the recession occurs, the forecasts will have an optimistic bias. Interviewees also pointed out that under-reporting crises and large recessions may be a response to the fear of triggering a worse, self-fulfilling crisis or recession.

---

[53] Controlling for the effect of important recessions is also motivated by the evidence—discussed since Mitchell (1927) and recently revisited by Morley and Piger (2012)—that recessions tend to occur more abruptly and be associated with temporary shocks, while expansions are more gradual and are frequently related to permanent shocks. These differences could surface as an asymmetry in GDP growth rates and have important implications for forecasting performance.

[54] In the context of short-term forecasts, Genberg and Martinez (2014b) show that the IMF's inability to predict recessions is not significantly worse than that of private sector forecasters. They also report that the optimistic biases disappear when recessions are removed from the sample, and suggest that inappropriate incentive structures maybe a factor in the poor forecasting of recessions. Also see Loungani (2001).

[55] The IMF's Early Warning Exercise, created in response to the recent financial crisis and jointly conducted with the G20's Financial Stability Board, aims at flagging low-probability events with high-impact risks to the global macroeconomic and financial stability. See Annex 5.

Forecast errors across area departments and levels of development

143. Table 7 summarizes the frequency of statistically significant biases in five-year-ahead forecasts of GDP growth, for countries grouped by IMF area department and different stages of economic development.[56] In most cases, only intervals of results are shown, based on the full set of results, including for the three- and four-year-ahead forecasts, which is reported in Tables A2.4 through A2.6, in Annex 2.

144. Regardless of the measure of bias (mean or median) and the method (descriptive or regression-based), statistically significant biases in medium-term forecasts of GDP growth seem evenly distributed across area departments.

145. Nevertheless, forecasts for countries in the African (AFR), Western Hemisphere (WHD), and European (EUR) departments, in that order, account for most of the biases.[57] Considering the entire IMF membership (roughly 180 countries), in which between 24 percent and 33 percent of countries show statistically significant biases, these departments combined account for between 16 percent and 27 percent of the cases.

Table 7. Frequency of statistically significant bias in 5-year-ahead forecasts across different country groupings using the country-by-country approach
(Percent of IMF member countries)

		Descriptive Statistics		Regression-based statistics
		Mean bias	Median bias	(Mean bias)
Full sample		27-31	30-33	24-33
Department	APD	4	4	3–6
	AFR	7–9	9–10	6–9
	EUR	4–7	6–8	4–6
	MCD	3–6	3–5	3–6
	WHD	5–9	6–9	4–9
Development	AE	3–6	4–6	4–6
	EME	11–14	12–14	9–14
	LIC	12–14	12–14	10–16
Other groupings	OECD	4–6	4–6	4–6
	G7	2	2–3	1–2
	G20	4	4–5	3–5
	OIL	2–3	2–3	2–4
	Primary Exp.	3–5	4–6	3–6

Source: Table A2.4, Annex 2.

Note: Only ranges of results are reported. Tests of unbiasedness rely on two-sided *t*-tests (mean) and Wilcoxon tests (median).

[56] The underlying statistical tests of unbiasedness follow the country-by-country approach using both descriptive and regression-based analyses.

[57] This does not necessarily mean that these departments have a higher propensity for biased forecasts, since results are not computed as a proportion of the number of countries in each area department.

146. Among country groupings based on level of development, less developed economies (both emerging-market (EME) and low-income (LIC) economies), account for most of the biases (between 19 percent and 30 percent). Countries covered by the Middle East and Central Asia (MCD) and Asia-Pacific (APD) departments, as well as advanced (AE), G7, and oil-exporting economies, are responsible for small shares of the biases.

147. Tables 8 and 9 summarize the average cross-country biases from the country-by-country and aggregate analyses, respectively. The full set of results for the aggregate analysis is reported in Tables A2.7.

148. On average, forecasts for countries in all regions and development stages display the optimistic bias found in the full sample, regardless of the measure of bias, the forecast method, the forecast horizon, *WEO* vintage (Spring or Fall), or how out-turns are measured (one-year or two-year delay window). For instance, cross-country average biases in five-year-ahead forecasts for countries in EUR range from -0.9 percentage points to -1.1 percentage points, when the simple descriptive mean is used to measure biases in individual countries, and from -0.1 percentage points to -0.3 percentage points when the median is used (Table 8).[58] Average optimistic biases are also found for economies in the OECD, in the G7 and G20 groupings, and for exporters of primary products.

149. The exception to the tendency towards optimistic biases is the group of countries in MCD, for which there is *underprediction* of GDP growth, on average, perhaps reflecting the heavy concentration of oil exporters in this region (for oil exporters, the average bias is also pessimistic).

150. The largest absolute bias seems to occur in forecasts by the EUR and WHD area departments, as well as for LICs, consistently with the larger frequency found in these same country groupings. However, somewhat surprisingly, forecasts for advanced economies also seem to show large average optimistic biases. Especially according to the aggregate analysis based on pooled regressions (Table 9), for which outliers are excluded, average biases in advanced economies actually seem larger than those in both emerging market economies and LICs. Less frequent but larger forecast biases in advanced economies suggest a concentration of large biases in few countries. Forecasts for emerging market economies, on the other hand, consistently show the smallest average optimistic biases.[59]

[58] When the mean bias in a given country is computed using a regression of forecast errors on a constant and on its own lagged values, the cross-country average bias in EUR is -1 percentage points to -1.3 percentage points. The difference is caused by the different number of number of observations used in each case.

[59] In Table 9, the differences relative to biases in emerging economies are always statistically significant in the case of advanced economies and, often, in the case of LICs.

Table 8. Average bias in medium-term forecasts of GDP growth across different country groupings using the country-by-country approach
(In percentage points)

		Mean				Median				Regression-based[1]			
		1 Yr window		2 Yr window		1 Yr window		2 Yr window		1 Yr window		2 Yr window	
		Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall
5-Year-Ahead	Full sample	-0.48	-0.56	-0.44	-0.45	-0.23	-0.24	-0.12	-0.14	-0.64	-0.69	-0.53	-0.59
Department	APD	-0.59	-0.54	-0.58	-0.54	-0.37	-0.36	-0.24	-0.28	-0.49	-0.36	-0.47	-0.25
	AFR	-0.07	-0.57	-0.33	-0.61	-0.37	-0.44	-0.27	-0.27	-0.72	-1.20	-0.92	-1.36
	EUR	-1.13	-1.04	-1.02	-0.90	-0.34	-0.29	-0.15	-0.14	-1.31	-1.20	-1.08	-0.97
	MCD	0.53	0.50	0.86	0.99	0.81	0.76	1.06	0.98	0.95	0.84	1.55	1.45
	WHD	-1.01	-0.89	-0.92	-0.91	-0.71	-0.69	-0.81	-0.83	-1.31	-1.08	-1.33	-1.28
Development	AE	-0.94	-0.87	-0.85	-0.76	-0.38	-0.39	-0.26	-0.27	-0.94	-0.87	-0.85	-0.76
	EME	-0.21	-0.27	-0.12	-0.03	0.05	0.10	0.20	0.24	-0.21	-0.27	-0.12	-0.03
	LIC	-0.54	-0.70	-0.59	-0.75	-0.46	-0.52	-0.40	-0.48	-0.54	-0.71	-0.58	-0.75
	OECD	-0.98	-0.89	-0.89	-0.82	-0.42	-0.39	-0.31	-0.32	-0.98	-0.89	-0.89	-0.82
	G7	-1.04	-1.07	-1.06	-1.06	-0.79	-0.89	-0.71	-0.82	-1.04	-1.07	-1.06	-1.06
	G20	-0.82	-0.84	-0.79	-0.75	-0.18	-0.22	-0.14	-0.11	-0.82	-0.84	-0.79	-0.75
	OIL	1.58	1.00	1.71	1.57	1.44	1.25	1.67	1.56	1.58	1.00	1.71	1.57
	Primary	-0.79	-0.94	-1.15	-1.24	-0.63	-0.72	-0.89	-0.89	-0.59	-0.54	-0.58	-0.54
4-Year-Ahead	Full sample	-0.67	-0.67	-0.62	-0.53	-0.34	-0.33	-0.21	-0.18	-0.79	-0.75	-0.69	-0.53
Department	APD	-0.57	-0.51	-0.55	-0.48	-0.33	-0.29	-0.19	-0.14	-0.75	-0.63	-0.78	-0.53
	AFR	-0.67	-0.78	-0.89	-0.73	-0.59	-0.70	-0.48	-0.53	-1.08	-1.27	-1.33	-1.03
	EUR	-1.07	-1.03	-0.96	-0.89	-0.35	-0.26	-0.19	-0.08	-1.11	-1.03	-0.94	-0.86
	MCD	0.14	0.08	0.48	0.59	0.44	0.39	0.67	0.70	0.44	0.60	1.03	1.24
	WHD	-1.00	-0.89	-0.90	-0.89	-0.69	-0.62	-0.67	-0.68	-1.19	-1.03	-1.07	-1.06
Development	AE	-0.83	-0.79	-0.73	-0.70	-0.38	-0.37	-0.28	-0.23	-0.83	-0.79	-0.73	-0.70
	EME	-0.38	-0.44	-0.29	-0.17	-0.10	-0.08	0.09	0.13	-0.38	-0.44	-0.29	-0.17
	LIC	-0.90	-0.85	-0.92	-0.83	-0.57	-0.59	-0.50	-0.49	-0.88	-0.85	-0.92	-0.83
	OECD	-0.90	-0.85	-0.81	-0.77	-0.36	-0.38	-0.30	-0.25	-0.90	-0.85	-0.81	-0.77
	G7	-1.03	-1.05	-1.05	-1.06	-0.78	-0.84	-0.79	-0.80	-1.03	-1.05	-1.05	-1.06
	G20	-0.81	-0.81	-0.77	-0.76	-0.22	-0.19	-0.09	-0.09	-0.81	-0.81	-0.77	-0.76
	OIL	0.66	0.59	0.72	1.20	0.60	0.52	0.98	0.95	0.66	0.59	0.72	1.20
	Primary	-0.97	-1.12	-1.28	-1.34	-0.44	-0.74	-0.76	-0.93	-0.57	-0.51	-0.55	-0.48
3-Year-Ahead	Full sample	-0.75	-0.76	-0.69	-0.64	-0.41	-0.37	-0.33	-0.24	-0.72	-0.65	-0.53	-0.45
Department	APD	-0.52	-0.43	-0.46	-0.37	-0.37	-0.21	-0.16	-0.08	-0.65	-0.53	-0.57	-0.49
	AFR	-0.74	-0.92	-0.93	-0.89	-0.71	-0.80	-0.74	-0.59	-0.67	-0.60	-0.65	-0.48
	EUR	-1.17	-1.11	-1.03	-0.98	-0.50	-0.37	-0.31	-0.17	-1.17	-1.10	-0.97	-0.88
	MCD	-0.14	-0.22	0.10	0.20	0.35	0.29	0.43	0.52	0.10	0.08	0.60	0.72
	WHD	-0.98	-0.91	-0.89	-0.91	-0.61	-0.54	-0.65	-0.67	-1.06	-0.91	-0.84	-0.86
Development	AE	-0.90	-0.85	-0.79	-0.76	-0.51	-0.44	-0.39	-0.30	-0.90	-0.85	-0.79	-0.76
	EME	-0.53	-0.56	-0.42	-0.33	-0.13	-0.09	-0.06	0.12	-0.53	-0.56	-0.42	-0.33
	LIC	-0.90	-0.94	-0.93	-0.92	-0.66	-0.63	-0.58	-0.59	-0.90	-0.93	-0.92	-0.92
	OECD	-0.97	-0.91	-0.87	-0.83	-0.51	-0.47	-0.43	-0.36	-0.97	-0.91	-0.87	-0.83
	G7	-1.16	-1.18	-1.18	-1.19	-0.91	-0.96	-0.92	-0.96	-1.16	-1.18	-1.18	-1.19
	G20	-0.86	-0.85	-0.80	-0.81	-0.18	-0.23	-0.16	-0.21	-0.86	-0.85	-0.80	-0.81
	OIL	0.39	0.14	0.45	0.62	0.32	0.35	0.34	0.76	0.39	0.14	0.45	0.62
	Primary	-1.17	-1.32	-1.39	-1.44	-0.70	-0.85	-0.80	-0.95	-0.52	-0.43	-0.46	-0.37

Source: Author's calculations using *WEO*.

(1) For each individual country, forecast errors are regressed on their own lagged values and a constant..

Table 9. Average bias in medium-term forecasts of GDP growth across different country groupings using the aggregate approach

(In percentage points)

		Spring *WEO*	Fall *WEO*
		5-year-ahead	
Department	APD	-0.33 (*)	-0.30 (*)
	AFR	-0.43 (*)	-0.46 (*)
	EUR	-0.44 (*)	-0.45 (*)
	MCD	0.56 (*)	0.57 (*)
	WHD	-0.80 (*)	-0.80 (*)
Development	AE	-0.66 (*)	-0.64 (*)
	EME	-0.18 (*)	-0.17
	LIC	-0.32 (*)	-0.36 (*)
		4-year-ahead	
Department	APD	-0.30 (*)	-0.22 (*)
	AFR	-0.60 (*)	-0.60 (*)
	EUR	-0.54 (*)	-0.47 (*)
	MCD	0.23	0.36 (*)
	WHD	-0.79 (*)	-0.79 (*)
Development	AE	-0.62 (*)	-0.61 (*)
	EME	-0.34 (*)	-0.25 (*)
	LIC	-0.45 (*)	-0.44 (*)
		3-year-ahead	
Department	APD	-0.32 (*)	-0.19
	AFR	-0.79 (*)	-0.77 (*)
	EUR	-0.56 (*)	-0.48 (*)
	MCD	0.05	0.10
	WHD	-0.82 (*)	-0.81 (*)
Development	AE	-0.68 (*)	-0.62 (*)
	EME	-0.37 (*)	-0.29 (*)
	LIC	-0.63 (*)	-0.59 (*)

Source: Table A2.7, Annex 2.
Note: (*) Indicates statistical significance at the 5% level in at least one specification (pooled least square or fixed-effect estimators).

151. Indeed, biases in G7 economies, a subset of advanced economies, are much less frequent and larger, on average, than biases in LICs, mainly because of biases for France, Germany, Italy, and Japan. Timmermann (2006) also flagged the large biases in these four G7 economies.[60]

152. The results in Tables 7–9, showing optimistic biases regardless of countries' development level are interesting when compared with country authorities' perceptions about IMF forecasts reported in Genberg and Martinez (2014a). The average optimistic biases for advanced economies are often the largest, and those for emerging market economies are

[60] Similar results to Timmermann's in this and other instances may be due to sample-related issues. Despite the seven-year lag between the two studies, some forecasts that became available after Timmermann's study are not included in this evaluation because no "actual GDP" data are yet available for computing forecast errors.

always the smallest. The survey data, however, indicate that authorities in advanced economies tend to perceive IMF growth forecasts for emerging markets as too high, while authorities in emerging market economies—more in line with the evidence presented in this section—perceive the same forecasts for their own economies as being too low and often view forecasts for advanced economies as too high.

153. Table 9 also shows that in all country groupings, with few exceptions, the measured biases are statistically significant, confirming the existence of biases in all area departments and in countries in different stages of development. These aggregate results, however, must be taken as a complement to the country-by-country figures presented in Tables 4 and 7, which show problematic cases to be concentrated in less than 30 percent of member-countries in a statistically significant sense.

154. Taken together, these results suggest that, although there are occurrences in all departments, biases are not pervasive, being limited to about one third of the membership. Nor is it associated with intrinsic problems in the forecasting approach at the institutional level, though a more appropriate structure of incentives may improve the performance of IMF forecasters, specifically in forecasting recessions.

155. However, the frequency of biases is not negligible either. IMF Management and the Board of Executive Directors should perhaps consider whether this problem deserves some corrective action.

Is the optimistic bias larger if a country is in an IMF program?

156. This subsection revisits one perception in the literature,[61] and among some country officials, that IMF forecasts made in the context of programs are more overoptimistic than those made in non-program situations.

157. Based on the interviews conducted by the evaluation team, the same perception exists among some members of the Executive Board and staff. In post-survey interviews with 50 randomly selected IMF staff, mostly desk economists, about half the respondents indicated that "other considerations which might also influence the forecasts in program countries" tend to induce optimistic biases.[62]

---

[61] See, for example, Timmermann (2006), Batista and Zalduendo (2004), IMF (2004), GAO (2003), and Heritage Foundation (1999).

[62] Additional optimistic biases in forecasts made during programs, often found in the literature, may be related to the way the IMF conducts forecasts in these situations. Because IMF projections during programs are *conditional forecasts*—made under the assumption that all the conditionality criteria and quantitative targets will be met, presumably improving the economic conditions, including GDP growth—the optimistic bias might be built in, since conditionality is not always respected. Luna (2014) analyzes forecasts during IMF programs and describes other reasons why there may be extra optimism in these forecasts.

158. As described in what follows, our analysis of medium-term forecasts for program cases confirms the general findings in the literature on short-term forecasts, but shows a more nuanced picture of program-related optimistic biases. Results are displayed in Tables 10–12.[63] Further details of the methodology and results are provided in Annex 3.

159. Considering errors in forecasts made at year *t* for year *t+k*, observations are first sorted according to whether the country was in an IMF program at year *t* or not.[64] We find that the mean forecast errors are always negative and statistically significant regardless of whether forecasts are made while the country is in an IMF program or not, suggesting that optimistic biases exist in both situations (Table 10). However, we find no statistically significant differences between the average bias in program and no-program cases (except in three-year-ahead forecasts, where the differences are only statistically significant for the Fall vintage). In both program and no-program cases, biases are close to the full sample averages shown in Table 5.[65] Solely based on results in Table 10, the evidence of any *additional* optimistic bias in IMF programs, beyond the optimism observed in general, is not very strong.

Table 10. IMF programs (I): bias in medium-term forecasts of GDP growth
(In percentage points)

	Spring *WEO*	Fall *WEO*
	5-year-ahead	
No program at year *t*	-0.33 (*)	-0.36 (*)
Program at year *t*	-0.34 (*)	-0.27 (*)
Statistically different?	No	No
	4-year-ahead	
No program at year *t*	-0.46 (*)	-0.37 (*)
Program at year *t*	-0.39 (*)	-0.45 (*)
Statistically different?	No	No
	3-year-ahead	
No program at year *t*	-0.50 (*)	-0.41 (*)
Program at year *t*	-0.62 (*)	-0.66 (*)
Statistically di*fferent?*	*No*	Yes

Source: Table A3.1, Annex 3.
Note: (*) Indicates statistical significance at the 5% level in at least one specification (pooled least square or fixed-effect estimators). See Annex 3.

160. Differences in forecast errors between program and no-program cases may be related to different treatment these economies receive from IMF forecasters, or driven by idiosyncratic factors that are outside the control of IMF forecasters, such as different types of shocks or policies affecting these two groups of economies. For example, the frequency of

[63] Figure A3.1 in Annex 3 provides a visual description of the statistically significant results in the baseline estimations (i.e., without controlling for "big recessions," as discussed below) for all vintages and horizons.

[64] In Figure A3.1, these two cases are represented in Boxes (B) and (C). See also the estimation of equation A3.1, in Annex 3.

[65] See also Figure A3.1, in Annex 3.

domestic shocks leading to recessions or the vulnerability to adverse global shocks may be higher in economies that seek IMF assistance more often.

161. To investigate the differences these possibilities, we next separate the observations into those for "program countries"—i.e., countries that were engaged in IMF programs at some point during the sample period—and those for countries with no history of programs over the same period—i.e., "non-program countries." Within the subsample for program-countries, observations can be further subdivided into those coming from program-countries that (i) are or (ii) are not in an IMF program at year *t*, when forecasts are made. The results are shown in Table 11.[66]

Table 11. IMF programs (II): bias in medium-term forecasts of GDP growth
(In percentage points)

	Spring *WEO*	Fall *WEO*
5-year-ahead		
"Non-program country"	-0.15 (*)	-0.16 (*)
"Program country"	-0.48 (*)	-0.47 (*)
Not in IMF program at year *t*	-0.61 (*)	-0.65 (*)
In IMF program at year *t*	-0.34 (*)	-0.27 (*)
4-year-ahead		
"Non-program country"	-0.18 (*)	-0.20 (*)
"Program country"	-0.65 (*)	-0.53 (*)
Not in IMF program at year *t*	-0.88 (*)	-0.61 (*)
In IMF program at year *t*	-0.39 (*)	-0.45 (*)
3-year-ahead		
"Non-program country"	-0.28 (*)	-0.30 (*)
"Program country"	-0.73 (*)	-0.60 (*)
Not in IMF program at year *t*	-0.82 (*)	-0.56 (*)
In IMF program at year *t*	-0.62 (*)	-0.66 (*)

Source: Tables A3.2 and A3.3, in Annex 3.

Note: (*) Indicates statistical significance at the 5% level in at least one specification (pooled least square or fixed-effect estimators). See Annex 3.

162. Results strongly suggest an additional optimistic bias in program countries, which is often more than twice as large as that in non-program countries; the difference is always statistically significant and is robust to different forecast horizons and vintages.[67] On average, forecasts for program countries tend to be more optimistic than forecasts for non-program countries by between 0.27 percentage points and 0.45 percentage points.

[66] In Figure A3.1, the results based on observations for non-program countries are shown in Box (D). Among program countries (whose aggregate results are in Box E), results for those not in an IMF program are in Box (F), while Box (G) shows results for countries engaged in a program at the time of the forecast.

[67] These findings are also robust to the introduction of dummy variables for important recessions.

163. Together, Tables 10 and 11 suggest that (i) forecast errors associated with program countries not in an IMF program at year *t* (when the forecast is made) show larger optimistic biases and that (ii) part of these biases may be driven not by the program status (at time *t*) itself but by factors that are specific to countries that more often seek IMF programs. One cannot be certain, however, whether those factors are outside the control of IMF forecasters.

164. When we divide the subsample of program countries into those that were and were not in an IMF program at time *t* (when the forecast is made), we find that regardless of whether or not program countries are actually in a program at year *t*, these countries always show larger optimistic biases than do non-program countries, and that these differences are often statistically significant. For example, considering the Spring vintage of five-year-ahead forecasts, the estimated bias in non-program countries is -0.15, while biases in program-countries are about four times as large (-0.61) when the country is not in a program and about twice as large (-0.34) when they are in IMF programs.[68]

165. Therefore, although forecasts made during programs are always more optimistically biased than forecasts for non-program countries, they are *less* optimistically biased than those in program countries that are not in a program at the time of the forecast. These differences are statistically significant in the case of four- and five-year-ahead forecasts, suggesting that, conditionally on being a program-country, IMF programs are associated with lower optimistic biases, although the causality is not clear.

166. It may be that within program countries, during a program, the optimistic bias, if any, will change from one stage of the program to the next, or may even pre-date the program and be carried over into the program years. Do programs themselves cause the reduction in the optimistic bias in program countries, or are there factors intrinsic to these economies, perhaps outside the control of the IMF, that raise the optimistic biases in forecasts made before or after programs?

167. To explore these questions we analyze the timing of optimistic biases in program countries around program dates. Forecast errors from program countries are classified according to the year that forecasts are made (year *t*) relative to the year when programs start (year T_0). The results, shown in Table 12, suggest that to a large extent the optimistic biases in program countries come from forecasts that are made before these countries enter IMF programs: the optimistic bias in forecasts made before T_0-3 (i.e., four years or more before

[68] Differences are always statistically significant when the comparison is between program countries not in a program at year *t* (the time of the forecast) and non-program countries (Figure A3.1, Boxes D and F). When comparing non-program countries (Box D) and program countries during a program (Box G), the larger bias observed during programs is only significant for three-year-ahead forecasts (both vintages) and four-year-ahead forecasts (Fall vintage).

program start) is sizeable—between -0.65 and -0.92—and statistically significant for both Fall and Spring forecasts and all forecast horizons.[69]

Table 12. IMF programs (III): Bias in medium-term forecasts of GDP growth
(In percentage points)

	5-year-ahead		4-year-ahead		3-year-ahead	
	Spring	Fall	Spring	Fall	Spring	Fall
"Non-program-country"	-0.15 (*)	-0.16 (*)	-0.18 (*)	-0.20 (*)	-0.28 (*)	-0.30 (*)
"Program-country"	-0.48 (*)	-0.47 (*)	-0.65 (*)	-0.53 (*)	-0.73 (*)	-0.60 (*)
before T_0-3	-0.82 (*)	-0.72 (*)	-0.89 (*)	-0.65 (*)	-0.92 (*)	-0.65 (*)
T_0-3	-0.32	-0.75 (*)	-0.32	-0.30	-0.58	-0.60
T_0-2	-0.70 (*)	-0.58 (**)	-0.80 (*)	-0.48	-0.93 (*)	-0.26
T_0-1	-0.10	-0.47 (**)	-1.08 (*)	-0.65 (*)	-0.51 (*)	-0.40 (**)
T_0 (program start)	-0.23	-0.40 (**)	-0.64 (*)	-0.95 (*)	-0.99 (*)	-0.77 (*)
T_0+1	-0.26	-0.06	-0.27	-0.21	-0.69 (*)	-0.95 (*)
T_0+2	-0.36	-0.29	-0.40 (**)	-0.27	-0.32	-0.52 (*)
T_0+3	-0.81 (*)	-0.53 (*)	-0.20	-0.12	-0.65 (*)	-0.21
after T_0+3	-0.26	0.08	0.23	0.79	1.10 (**)	1.47 (**)

Source: Table A3.4, Annex 3.
Note: (*) and (**) indicate statistical significance at the 5% and 10% levels, respectively, in at least one specification (pooled least square or fixed-effect estimators). See Annex 3.

168. The results also show that the additional optimism found in IMF program cases versus non-program cases mainly arises from the forecasts made in the year that programs start. The overoptimism in these initial-year forecasts, which is perhaps influenced by considerations that are more binding at program inception, is usually large, ranging between -0.4 percentage point and -1 percentage point. Biases in forecasts made at T_0 are also always *larger* than the average bias for non-program countries (Table 12 and Figure A3.1, Box D), although results are mixed when the comparison is made relative to program countries not in a program at time *t*. in Any case, these larger optimistic biases at the start of programs typically disappear within one year after the program starts.

169. We conclude from this analysis that the evidence of additional optimism in medium-term GDP growth projections related to IMF programs is stronger for shorter forecast horizons and seems to occur mainly in the year the program starts, being quickly corrected after that.[70]

[69] In Figure A3.1, in Annex 3, this case is shown in the first box to the left of Box (H).

[70] Using the Monitoring of Fund Arrangements (MONA) database, Luna (2014) shows that optimistic biases in short-term forecasts of GDP growth observed at the inception of programs tend to be corrected at the time of the first program review. He also finds that other variables more closely related to quantitative targets—such as fiscal and current account balances—show a rather conservative bias (lower deficits than projected) that persists in subsequent reviews. Based on evidence from both the IEO evaluation survey and interviews with IMF desk economists, Luna suggests that considerations regarding the "selling" of the program may explain the optimism in GDP projections, on the one hand, while incentives to make the program "viable" (i.e., complete all the scheduled reviews), by setting easier targets to be met, may explain the rather conservative bias in other

(continued…)

Possible reasons for biases in program cases

170. Ambiguity about the causes of program-related biases was reflected in the interviews that the evaluation team conducted with IMF staff. While about half the respondents agreed that "other considerations which might also influence the forecasts in program countries" tend to induce an optimistic bias, as previously reported, nearly as many respondents (42 percent) indicated that "it depends."

171. Among the factors that could affect biases during programs, staff (mostly desk economists) mentioned the relative strength of technical arguments and political positions from, on one side, the country authorities, who tend to push for more optimistic forecasts of GDP and, on the other side, the IMF, which could sometimes tend to be more conservative and bias the projections in the pessimistic direction. In any case, the interviews seem to confirm the notion that forecasts are sometimes a result of a compromise, a negotiation,[71] between the IMF and country authorities and that this compromise may have an ambiguous effect on the forecasts made during programs, pushing them either in the optimistic or pessimistic direction. Staff also mentioned that, after a program starts, forecasts receive greater scrutiny and Fund economists often have access to better data; they suggested that these factors, too, may improve the quality of forecasts.

172. There are several possible reasons (not mutually exclusive) for the optimism in GDP growth projections when programs start. First, as discussed earlier (footnote 62), the optimism may be a consequence of the conditional aspect of forecasts made during programs—they assume that program targets will be met (which could help on the growth front), but is not always the case.

173. Second, optimistic biases during programs may be driven by factors that are specific to countries that seek IMF programs, not by the program status itself. For example, if GDP growth in program countries is typically reduced some time after programs are completed, the optimism in medium-term GDP forecasts may be simply a result of the already discussed inability to predict decelerations in growth.

174. This possibility is illustrated with the help of Figure 8. Panel (a) shows the profile of GDP growth in a typical program country around program dates.[72] Note that the annual GDP growth rate falls from about 3.6 percent—the sample average—to about 2 percent at year T_0-3 (three years before program start). The statistically significant bias in forecasts made

---

variables, on the other hand. Once the programs start, the selling aspect is no longer an issue, the optimistic biases in high-profile variables, such as inflation and GDP, disappear, and only the conservative bias in target variables remains.

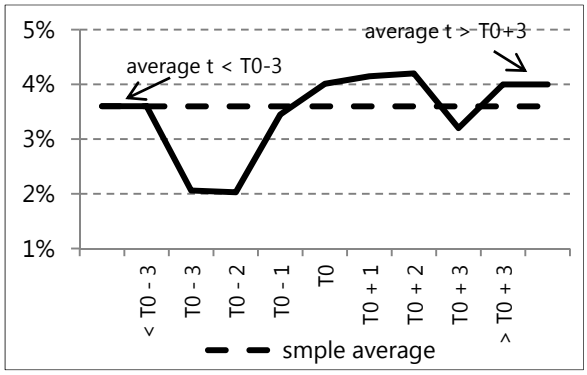
[71] "Negotiation" here does not necessarily imply any *quid pro quo*.

[72] The average growth is computed from observations from program countries in the sample.

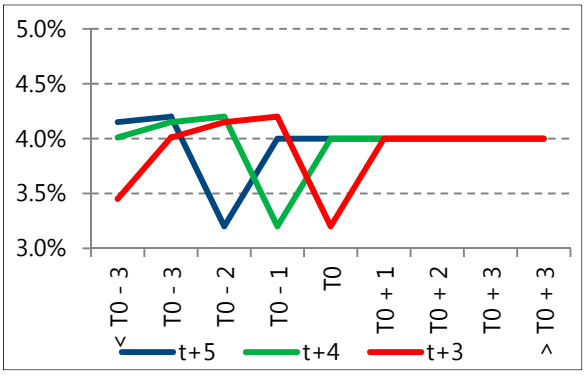
prior to T_0-3 could thus largely reflect the inability of IMF forecasters to predict the deceleration in growth typically observed before the start of IMF programs.

Figure 8. Timing of forecast biases in programs

(a) Typical Program country: Actual GDP growth at *t*



(b) Correct Forecasts made at time *t* (i.e., actual GDP growth at *t* + *k*)



Source: Author's calculations using the Spring vintages of *WEO*.

175. After falling at T_0-3, GDP growth in a typical program country rises above the sample average rate between years T_0-1 and T_0+2, but drops again in T_0+3 (to about 3 percent annually), a time when most programs in the sample have already ended. The biases in forecasts made from T_0-2 to T_0 may be reflecting this second, usually post-program, deceleration in growth among program countries.

176. If the profile for GDP growth in panel (a) were accurately forecast, the optimal *k*-year-ahead forecast at year *t* would look exactly like the actual GDP growth observed in year *t*+*k*. Panel (b) in Figure 8 shows these optimal forecasts for the three horizons. If forecasters are unable to predict large decelerations in growth, panel (b) is very illustrative about the pattern of statistically significant biases during programs shown in Table 12 above.[73]

177. Consider, for example, forecasters making five-year-ahead forecasts at time T_0-2, two years prior to program start. They would need to predict the drop in GDP growth that is typically observed at T_0+3. If they miss that, an optimistic error will follow. The forecasters will face the same challenge when making four-year-ahead forecasts in the following year (T_0-1) and three-year-ahead forecasts in the year after that (T_0). This may explain why three-year-ahead forecasts show stronger evidence of a program-related additional bias in general, and at the time programs start, in particular, while five-year-ahead forecasts show little: the deceleration in growth observed at T_0+3 in a typical program country falls right into the three-year-ahead forecasts at the time of program start, but only affects five-year-ahead

[73] And, in Annex 3, Figure A3.1's green boxes around Box (H).

forecasts made two years before program start. It may also explain why four-year-ahead forecast biases are at their largest one year before program start (Figure 8).

178. We conclude that given the inability to predict sudden decelerations in growth, the observed pattern for optimistic forecast biases around program dates can be explained, at least partly, by the typical profile of GDP growth among program countries in the sample.

C. Can Other Variables Help Explain Forecast Errors?

179. Besides unbiasedness, when all available information is used, so that forecasts are efficient, forecast errors must not be "predicted" or explained by any variable available to the forecaster at the time of the forecast. Here we look at two potential aspects of informational inefficiency in IMF medium-term forecasts—serial correlation and failure to take economic interdependencies into account—and one informal rule used by Fund economists that may possibly create biases. Annex 2 parts C and D provide further details.

Serial correlation

180. If forecasts are efficient, past forecast errors should not help explain current errors. That is, forecast errors should not be serially correlated. Using both descriptive statistics and regress*i*on-based statistics, we find that serial correlation of forecast errors may be a problem only for fewer than about 20 percent of countries (Table 13).[74] As in the case of unbiasedness, the results indicate that when small-sample distortions are accounted for, serial correlation may be even less frequent than suggested by standard tests. This low incidence suggests the IMF has little scope to use the predictive power of past errors to help improve its current forecasts.

Table 13. Frequency of serial correlation in IMF medium-term forecasts of GDP growth
(Percent of countries)

	5-year-ahead	4-year-ahead	3-year-ahead
Descriptive statistics	14–22	8–11	7–9
Regression-based statistics	16–20	10–13	9–11

Source: Tables A2.1 and A2.2, Annex 2.

[74] The tests based on descriptive statistics simply assess whether the correlation between forecast errors and their *k*th-order lag is statistically different from zero. The regression-based results derive from the regression of errors on a constant, which captures the mean error, and on lagged errors, which captures the serial correlation. Table 13 reports the frequency at which the coefficient on the lagged error is statistically different from zero when serial correlation is tested independently from unbiasedness (zero mean error). The joint test—that both the constant and the autocorrelation coefficient are zero— is more restrictive and makes it easier to find evidence of serial correlation (around 40 percent of countries; see Annex 2).

Economic interdependencies

181. More generally, informational efficiency in forecasts requires that information on any variable x_t that is available at the time of the forecast must not help explain forecast errors. Following Timmermann (2006), this hypothesis is tested by estimating a regression of forecast errors on x_t. Efficiency is rejected if the estimated coefficient associated with x_t is statistically different from zero.

182. Timmermann (2006) reports that forecasts for major economies such as the United States and Germany can help predict IMF forecast errors and, therefore, can be used to improve the accuracy of these forecasts. Genberg and Martinez (2014a) update Timmerman's analysis and show that, although such economic interdependencies are being taken into account, this problem may still be an issue for short-term forecasts.

183. Table 14 shows the results for four variables taking the role of x_t: for k =3, 4, 5, the k-year-ahead forecasts of GDP growth for three large economies (Unites States, Germany, and China), and of oil prices. Overall, forecasts of these variables are correlated with errors in medium-term forecasts of GDP growth in a notable number of countries. Forecasts of U.S. GDP growth can explain errors in all IMF area departments, but they are especially important for countries covered by the Western Hemisphere Department (WHD), where they explain errors for more than 30 percent of the countries. U.S. forecasts are least important for countries covered by the IMF's European Department (EUR), where forecasts for Germany's GDP growth are the most important (explaining errors for as many as 20–30 percent of countries, in three- and four-year-ahead forecasts). Forecasts of China's GDP growth seem to be important for errors in countries covered by Middle East and Central Asian Department (MCD), and forecasts of oil prices explain a large portion of the errors in countries in the African Department (AFR).

184. The fact that forecast errors in GDP growth can be reduced, in principle, by taking into account forecasts for other countries and for oil prices indicates that international interdependencies may have not been fully taken into account in IMF medium-term forecasts.

185. By comparing our results (Table 14) with those obtained by Genberg and Martinez (2014b), we find that this type of informational inefficiency—linked to taking insufficient account of international interdependencies and spillovers—is more frequent in IMF medium-term than short-term forecasts. While here the frequency of signs of informational inefficiency are almost always higher than 10 percent and often in the

20−30 percent interval, Genberg and Martinez (2014b) report frequencies usually not much greater than 5 percent.[75]

Table 14. Informational inefficiency in IMF medium-term forecasts of GDP growth
(Share of countries)

	Spring				Fall			
	US	Germany	China	OIL	US	Germany	China	OIL
	3-year forecast errors							
AFR	0.20	0.14	0.18	0.30	0.18	0.14	0.16	0.23
APD	0.23	0.13	0.17	0.13	0.20	0.10	0.17	0.13
EUR	0.15	0.20	0.10	0.10	0.08	0.28	0.08	0.05
MCD	0.20	0.13	0.27	0.07	0.17	0.10	0.10	0.10
WHD	0.35	0.29	0.21	0.09	0.38	0.26	0.06	0.18
	4-year forecast errors							
AFR	0.16	0.16	0.14	0.14	0.27	0.14	0.14	0.20
APD	0.20	0.07	0.10	0.17	0.27	0.10	0.10	0.17
EUR	0.10	0.33	0.08	0.20	0.10	0.20	0.03	0.05
MCD	0.20	0.20	0.27	0.17	0.17	0.13	0.23	0.10
WHD	0.32	0.26	0.12	0.12	0.35	0.18	0.06	0.12
	5-year forecast errors							
AFR	0.20	0.20	0.20	0.23	0.27	0.27	0.25	0.20
APD	0.23	0.17	0.13	0.13	0.30	0.17	0.17	0.13
EUR	0.10	0.15	0.10	0.10	0.10	0.13	0.10	0.10
MCD	0.20	0.17	0.23	0.17	0.10	0.10	0.17	0.10
WHD	0.35	0.18	0.18	0.18	0.38	0.09	0.15	0.06

Source: Author's calculations using *WEO*.

The "Five-Year-Rule" for the closing of output gaps

186. Another possible source of optimistic biases in IMF medium-term forecasts of GDP growth, previously discussed by Timmerman (2006), may be the use of an informal rule that imposes the closing of the output gap within five years. Based on post-survey interviews with staff, the mandatory use of this rule has been relaxed after the 2008–09 financial crisis, but the rule continues to be used by Fund economists.

187. Desk economists who support its use argue that reasons for the rule to be violated in practice are rare events,[76] and that the rule provides a useful guide to economists, especially when no quantitative forecasting model is available and judgment is the main forecast method.

[75] Genberg and Martinez (2014b) also show that a measure of global interdependence in (actual) GDP growth proposed by Matheson (2013) is, to a large extent, also found in IMF short-term forecasts of GDP growth, indicating that these forecasts are taking into account, although not fully, interdependencies between economies.

[76] Several interviewees mentioned, as an example, the recovery from the 2009 crisis, which has been unusually slow by historical standards.

188. Using such an arbitrary rule may nonetheless introduce distortions in forecasts. Timmerman (2006), in his analysis of short-term forecasts, argues that if the rule induces desk economists to assume a premature closing of the output gap, not observed in reality, the resulting forecasts will tend to overpredict GDP growth. In that case, when economists try to respect the informal rule, smaller predicted output gaps (in absolute terms) for year $t+k$ should be associated with larger absolute forecast errors for that year. The correlation between errors and the *predicted* output gap should therefore be negative.

189. Timmermann's results are confirmed by those in Table 15, suggesting that a premature closing of the output gap may be leading to overpredictions of GDP growth also in medium-term forecasts. The table shows the correlations for both the Spring and Fall forecast vintages, with out-turns measured with delays of one or two years. The correlations range from about -0.1 to -0.3 and the hypothesis that they are equal to zero is strongly rejected in all cases.

Table 15. Correlations between forecast errors and forecasts of the output gap at $t+k$

	Spring		Fall	
	1 Year	2 Year	1 Year	2 Year
	5-Year-Ahead			
full sample				
Correlation	-0.14	-0.16	-0.21	-0.22
p-value	0.00	0.00	0.00	0.00
Advanced				
Correlation	-0.14	-0.12	-0.18	-0.17
p-value	0.00	0.01	0.00	0.00
G7				
Correlation	-0.19	-0.23	-0.25	-0.25
p-value	0.05	0.02	0.01	0.01
	4-Year-Ahead			
full sample				
Correlation	-0.12	-0.14	-0.21	-0.22
p-value	0.00	0.00	0.00	0.00
Advanced				
Correlation	-0.12	-0.10	-0.17	-0.16
p-value	0.01	0.02	0.00	0.00
G7				
Correlation	-0.21	-0.24	-0.28	-0.28
p-value	0.03	0.01	0.00	0.00
	3-Year-Ahead			
full sample				
Correlation	-0.12	-0.15	-0.20	-0.21
p-value	0.00	0.00	0.00	0.00
Advanced				
Correlation	-0.12	-0.10	-0.16	-0.15
p-value	0.01	0.02	0.00	0.00
G7				
Correlation	-0.20	-0.24	-0.28	-0.27
p-value	0.04	0.01	0.00	0.00

Source: Author's calculations using *WEO*.

190. Timmermann also argues that distortions caused by the informal rule on the closing of the output gap may be more serious in G7 countries. Indeed, Table 15 shows that, relative to the full sample, the problem is less severe in advanced economies as a group but is more prominent in G7 countries, whose results show larger negative correlations.

191. Potentially, then, the use of the informal five-year closing rule could be a factor in the biases observed in the medium-term forecasts for G7 economies, which are always larger than those in the full sample (Table 8), especially because of the large optimistic biases for Germany, France, Italy, and Japan.[77]

192. Because our sample does not contain forecasts made after 2008, these results may not capture changes that may have taken place in the application of the informal rule in response to Timmermann's recommendation that the rule be abandoned.

D. Are IMF Medium-Term Forecasts of GDP Growth Accurate?

193. Section II reported that forecasts based on mechanical methods for the estimation of potential output are positively correlated with IMF medium-term forecasts for GDP growth, and that the correlations (with the exception of the case of a linear trend) are statistically significant, albeit not very strong.

194. This finding, combined with survey results that show a widespread use of judgment in forecasts by IMF desk economists, indicate that the process of generating medium-term forecast is more complex, not necessarily in terms of sophistication of quantitative methods used, but in the sense of encompassing country-specific knowledge of economists and mission chiefs involved in the forecasts.

195. In light of this, how do IMF medium-term forecasts compare with forecasts based on naïve methods? How do they compare with forecasts from the private sector?

Comparisons with naïve and short-term forecasts

196. We compare the accuracy of actual IMF medium-term forecasts relative to medium-term forecasts resulting from naïve methods and to IMF short-term forecasts (those for the current and next year), as shown in Table 16, using ratios of root-mean-square errors (RMSE).[78] If the ratio RMSE(A)/ RMSE(B) is sufficiently below one (such that the

[77] Some stakeholders, such as some country officials who were interviewed by the evaluation team, were aware of the optimistic biases in these large European economies and in Japan, and attributed this outcome to political influence exerted by these economies on the IMF, although no objective evidence has been provided.

[78] For the forecasts based on the HP filter, two alternative values for the smoothing parameter (6.5 and 100) are considered (see discussion in Section II and Annex 1).

difference in forecasts is considered statistically significant), forecast A is judged more accurate than B.

Table 16. Root square mean errors in *WEO* forecasts of GDP growth

	5-Year Ahead				4-Year Ahead				3-Year Ahead			
	1 Yr window		2 Yr window		1 Yr window		2 Yr window		1 Yr window		2 Yr window	
Year	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall
WEO	0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Ratio to other naïve methods:												
linear	0.84	0.84	0.82	0.83	0.87	0.86	0.85	0.84	0.85	0.85	0.85	0.83
RW	0.86	0.29	0.84	0.85	0.89	0.30	0.87	0.86	0.88	0.29	0.87	0.86
HP (6.5)	0.67	0.66	0.63	0.65	0.71	0.70	0.69	0.68	0.72	0.71	0.70	0.69
HP (100)	0.72	0.72	0.68	0.70	0.76	0.75	0.73	0.73	0.76	0.76	0.75	0.74
MA(5)	0.74	0.74	0.71	0.72	0.80	0.78	0.78	0.77	0.77	0.76	0.76	0.75
Ratio to short-term forecasts:												
current year	1.82	2.56	1.93	2.65	1.82	2.54	1.93	2.63	1.79	2.47	1.91	2.56
% countries < 1	7.2%	4.4%	4.4%	2.8%	7.2%	3.9%	2.8%	2.8%	4.4%	4.4%	2.2%	3.8%
next year	1.12	1.23	1.21	1.29	1.13	1.21	1.21	1.27	1.10	1.15	1.19	1.21
% countries < 1	40.9%	23.8%	28.7%	19.9%	39.8%	24.9%	29.3%	20.4%	37.6%	20.9%	28.7%	18.7%

Source: Author's calculations using *WEO*.

197. We find that *WEO* forecast errors are clearly smaller than the errors produced by naïve forecasts, as indicated by RSME ratios that are always less than one. As expected, medium-term forecasts are much less accurate than current-year forecasts (about half as accurate, on average), or next-year forecasts (between 10 percent and 30 percent less accurate). Odd situations in which medium-term forecasts are more accurate than current-year forecasts occur in fewer than 7 percent of countries. But among next-year forecasts such situations are seen in as many as 41 percent of countries—which is troubling. Also, as expected, the average accuracy of medium-term forecasts tends to be greater for shorter horizons.

Comparison with consensus forecasts

198. How do IMF medium-term forecasts of GDP growth compare with forecasts from the private sector in terms of accuracy? Table 17 compares the *WEO* forecasts with those published by Consensus Economics. The latter represents forecasts from a pool of forecasters and it is released at about the same time as the *WEO.* Consensus publishes medium-term forecasts of GDP growth for countries covered by the IMF's Asia and Pacific (APD), European (EUR), and Western Hemisphere (WHD) departments.[79] For both the *WEO* and

[79] Consensus Economics publishes medium- and long-term (six to ten years ahead) forecasts twice a year, in April and October, for most of the economies it covers, and in March and September, for Eastern European economies. The economies considered in the comparisons here are: Argentina, Australia, Brazil, Canada, Chile, China, Colombia, Czech Republic, Euro Area, France, Germany, Hong Kong, Hungary, India, Indonesia, Italy, Japan, Korea, Malaysia, Mexico, New Zealand, the Netherlands, Norway, Peru, Poland, Romania, Russia,

(continued…)

Consensus, we calculate forecast errors using actual GDP growth rates measured two years after the target date.

Table 17. Ratio of RMSE (*Consensus*/*WEO*)

	Spring, 5-Year-Ahead Forecasts				Fall, 5-Year-Ahead Forecasts			
	>1		<1		>1		<1	
	Total	Stat. sign.	Total	Stat. sign.	Total	Stat. sign.	Total	Stat. sign.
APD	5	2	7	3	2	1	10	4
EUR	8	4	10	5	9	3	9	3
WHD	6	1	3	0	3	1	6	1
TOTAL	19	7	20	8	14	5	25	8
	Spring, 4-Year-Ahead Forecasts				Fall, 4-Year-Ahead Forecasts			
	>1		<1		>1		<1	
	Total	Stat. sign.	Total	Stat. sign.	Total	Stat. sign.	Total	Stat. sign.
APD	4	0	8	2	3	0	9	0
EUR	9	2	9	2	8	2	10	1
WHD	3	0	6	1	3	0	6	1
TOTAL	16	2	23	5	14	2	25	2
	Spring, 3-Year-Ahead Forecasts				Fall, 3-Year-Ahead Forecasts			
	>1		<1		>1		<1	
	Total	Stat. sign.	Total	Stat. sign.	Total	Stat. sign.	Total	Stat. sign.
APD	5	0	7	2	6	0	6	1
EUR	6	2	12	3	7	3	11	3
WHD	4	0	5	0	4	0	5	0
TOTAL	15	2	24	5	17	3	22	4

Source: Author's calculations using *WEO* and Consensus Economics; statistical significance based on the Diebold-Mariano statistic at the 5 percent level.

199. Though overall we see that the Consensus forecasts tend to be more accurate than the IMF's (because the frequency of countries for which the RMSE ratio (Consensus/*WEO*) is less than one is larger than that of countries for which it is greater than one), the difference in performance is not substantial when considering only the statistically significant cases.[80] The difference seems to be driven mainly by problems in Asian and Pacific countries, where RMSE ratios that are less than one most clearly outnumber cases where the ratios are greater than one.

200. The worse performance of IMF economists relative to Consensus may reflect the more skeptical views of IMF desk economists regarding the reliability of medium-term

Singapore, Slovak Republic, Spain, Sweden, Switzerland, Taiwan, Thailand, Turkey, Ukraine, United Kingdom, United States, and Venezuela. See http://www.consensuseconomics.com. As discussed in Genberg and Martinez (2014b), "fair" comparisons of forecasts made by different institutions critically depend on whether the two forecasts are made using the same information set, which in turn depends on the relative lengths of forecast process leading up to publication. In the case of medium-term forecasts an informational advantage of a month may be less damaging to the fairness of comparisons than in the case of current- and next-year forecasts.

[80] The superior performance of private sector forecasts may be related to the fact that Consensus Economics averages across many forecasters.

forecasts for policy analysis relative to private sector forecasters and country officials. That perception—which may primarily capture the views of desk economists, and not those of IMF staff more generally or Management, given the institutional efforts to boost medium-term analyses in IMF products—emerges from the survey conducted by the evaluation team with desk economists, country authorities, and representatives from the private sector (global financial institutions), which is discussed in the Section VI.[81]

E. Summary of Findings on Forecast Performance

201. Findings on the performance of IMF medium-term growth forecasts can be summarized as follows:

- IMF medium-term point forecasts tend to overpredict GDP growth. Between two-third and three-fourth of the IMF membership show predicted growth rates, on average, higher than out-turns. In about 20–30 percent of countries this optimistic bias is statistically significant, but the frequency may be substantially lower if small-sample distortions are accounted for.

- The average optimistic bias ranges between 0.14 percentage points and 0.76 percentage points, depending on the forecast horizon, the measure (median or mean), and the method (descriptive or regression-based statistics, from either country-by-country or aggregate approaches).

- The optimistic bias is seen across different IMF area departments—with the notable exceptions of the Middle East and Central Asia (MCD) Department and oil exporters—and across different levels of economic development.

- Compared with the average, optimistic biases are less frequent but seem larger in advanced economies. Optimistic biases are more frequent in low-income and emerging-market economies, though the biases in the latter group are also the smallest.

- Serial correlation is less frequent (statistically significant in only 7 percent to 20 percent of countries) than bias, implying that the scope to use past forecast errors to improve forecasts across the IMF membership is limited.

- In 7 percent to 38 percent of countries, there are signs of general information inefficiency in forecasts: forecasts of GDP growth in large economies (United States,

[81] In their analysis of IMF short-term forecasts Genberg and Martinez (2014b) find that the accuracy of Consensus is *not* significantly different from that of the IMF. These different set of results could also originate in the IMF's less comprehensive approach to medium-term relative to short-tem forecasts as discussed in Section IV.

Germany, or China) or forecasts of oil prices can explain errors in medium-term forecasts of GDP growth. This type of informational inefficiency in forecasts is much more frequent in medium-term than in short-term forecasts, possibly indicating that efforts to account for international spillovers and interconnections have been implemented to a substantial degree in the latter but not in the former.

- To a great extent, biases in IMF medium-term forecasts of GDP growth seem to reflect an inability to predict crises and recessions. When the analysis controls for three recessions with potentially significant effects on large parts of the global economy (1997, 2001, and 2009) the measured optimistic biases tend to be eliminated, reversed, or substantially reduced.

- The use of an informal rule for the closing of the output gap, which may be forcing the forecasts of actual GDP to converge prematurely to the estimated level of potential output, may be another factor inducing optimistic biases in medium-term forecasts.

- Based on data from IMF programs over 1990–2012, there is no simple answer to the question of whether medium-term forecasts of GDP growth made during programs contain additional optimistic biases relative to non-program situations. The answer depends on:

    - The forecast horizon: strong evidence is only found in three-year-ahead forecasts.

    - The history of countries' participation in IMF programs: forecasts for countries with a history of IMF programs tend to be more optimistic than forecasts for non-program countries by between 0.27 percentage points and 0.45 percentage points.

    - The different stages of a program: large optimistic biases in medium-term forecasts of GDP growth pre-date the start of programs—perhaps reflecting the inability to predict the "crisis" that motivated the program. The biases resurface in the year that a program starts—perhaps reflecting considerations associated with the "selling" of the program at inception and/or the inability to predict a typical post-program deceleration in growth. But they typically disappear within one year after the program starts.

- IMF medium-term forecasts of GDP growth are clearly more accurate than those based on naïve forecast methods. Naïve methods produce forecasts that are mildly positively correlated with IMF forecasts, indicating that they may be part of the toolkit of IMF desk economists. Relative to Consensus forecasts, IMF forecasts are less accurate.

VI. How Users and Staff Perceive IMF Medium-Term Forecasts: Evidence from Survey Data

202. This section discusses how the IMF's medium-term growth forecasts are perceived by country authorities and the private sector and by the IMF desk economists responsible for making the forecasts. The evidence comes from the survey that the IEO undertook for the IMF Forecasts evaluation. Respondents from 179 country authorities, 126 IMF desk economists, and 26 representatives from internationally active financial institutions in the private sector participated in the survey.[82]

203. Next, in subsection A, we describe the responses of user groups and staff to the multi-part survey question specifically related to IMF medium-term forecasts, covering different aspects of medium-term forecasting activities at the Fund, including the importance of these forecasts and some aspects of their construction. Figure 9 shows the "response scores" from IMF desk economists and country authorities,[83] and Figure 10 compares the answers from desk economists, country authorities, and representatives from the private sector. Subsection B summarizes the main findings. Annex 4 provides detailed statistical comparisons of the responses from authorities and staff, as well as across different groups of countries, for each component of the medium-term survey question.

A. Responses to the Question on Medium-term Forecasting[84]

Staff, country authorities, and representatives of the private sector largely think that the medium-term forecasts for "their" countries are based on reasonable underlying assumptions about the global economy.

204. About four in five respondents across all three groups either agree or strongly agree with the Fund's assumptions underlying its medium-term forecasts (see prompt (a) in Figures 9 and 10). Although staff and authorities' answers are both strongly skewed to the left, suggesting the overall agreement with the statement, their average scores are not statistically different, suggesting that there is little evidence of a organizational bias in the desk economists' responses.[85]

---

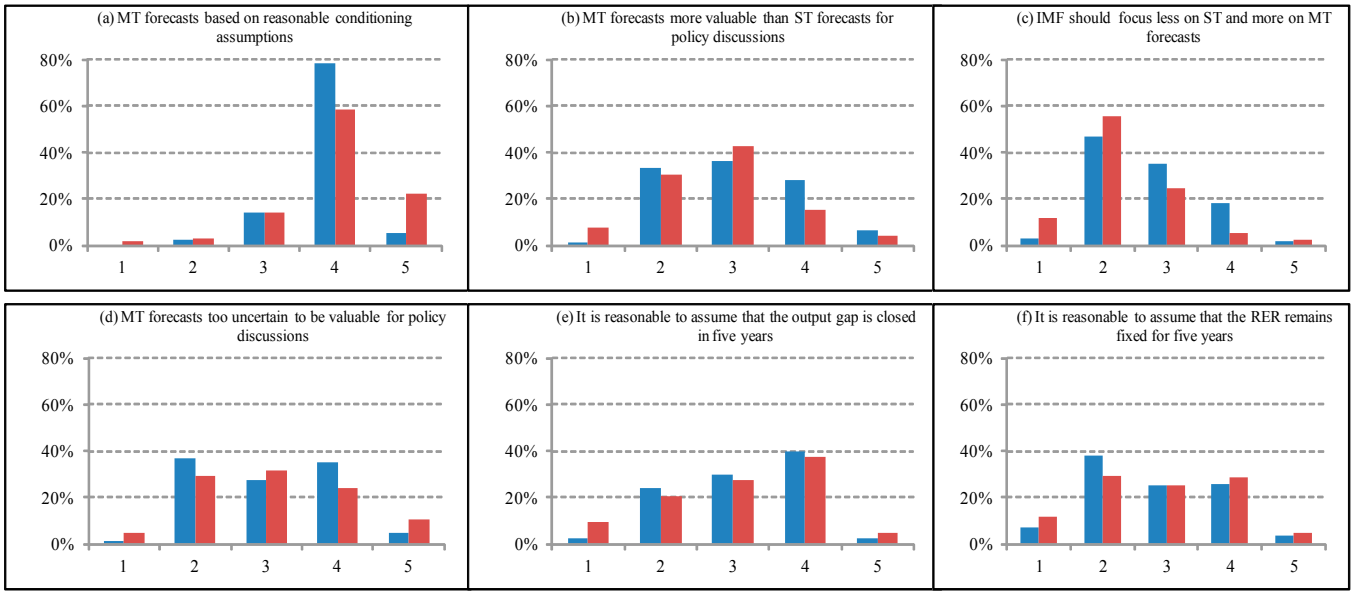
[82] For full details of the survey and survey responses, see Genberg and Martinez (2014a).

[83] The response scores are constructed from a scale that attaches values 1 to 5 to answers that, respectively, "strongly disagree," "disagree," "neither agree nor disagree," "agree," or "strongly agree" with prompts presented to participants. See Annex 4 for details.

[84] It should be emphasized that the survey question broadly refers to "medium-term forecasts" in general, without mentioning any specific variable or the type of forecast (point forecasts or scenarios and risk assessments).

[85] Annex 4 shows that the strong support, from both IMF desk economists and country authorities, to the underlying assumptions in medium-term IMF forecasts is robust to different partitions of the data, between program and non-program countries (Table A4.2), different levels of development (Table A4.3), or between different area departments (Tables A4.4 and A4.5(a)–(b)).

Figure 9. IMF desk economists and authorities' answers to survey question about IMF medium-term forecasts
(relative frequency, %)



Legend:
1 = "Strongly disagree"
2 = "Disagree"
3 = "Neither agree nor disagree"
4 = "Agree"
5 = "Strongly Agree"

Legend:
Authorities
IMF Desk Economists

Source: Author's calculations using IEO Forecast evaluation survey.

Questions:

(a) The IMF medium-term forecasts for my country are based on reasonable conditioning assumptions about world economic conditions.

(b) For policy discussions medium-term forecasts are more valuable than one or two-year forecasts.

(c) In its analysis, the IMF should place less enphasis on one and two-year forecasts and more enphasis on medium-term forecasts.

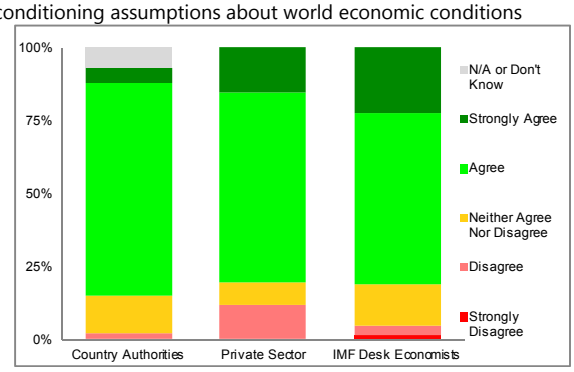
(d) medium-term forecasts entail too much uncertainty to be valuable for policy discussions.

(e) When making medium-term forecasts, its is reasonable to assume that my country's output gap is closed by the end of five years.

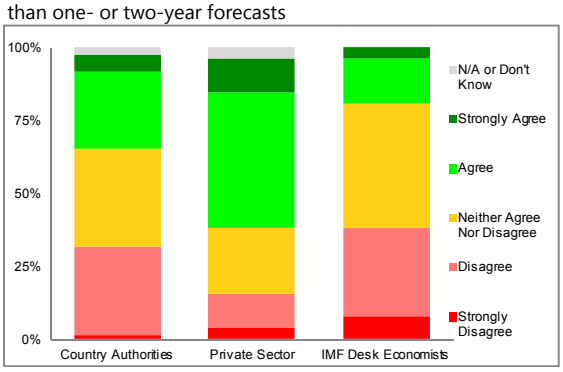
(f) When making medium-term forecasts, its is reasonable to assume that my country's real exchange rate remains fixed for five years.

Figure 10. Survey question: "Please rate your agreement with each of the statements below about the medium-term forecasts produced by the IMF."

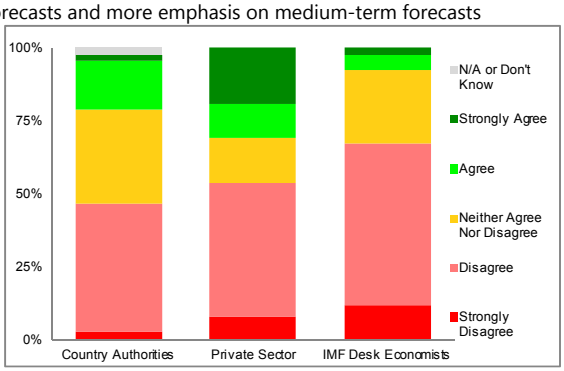
(a) The IMF medium-term forecasts for my country are based on reasonable conditioning assumptions about world economic conditions



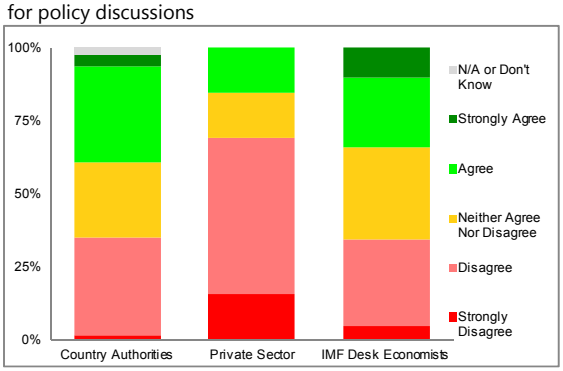
(b) For policy discussions medium-term forecasts are more valuable than one- or two-year forecasts



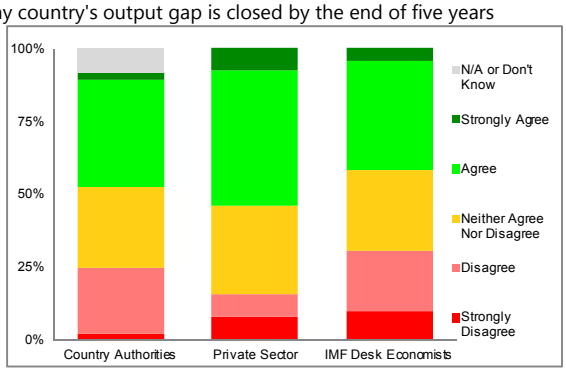
(c) In its analysis, IMF should place less emphasis on one- and two-year forecasts and more emphasis on medium-term forecasts



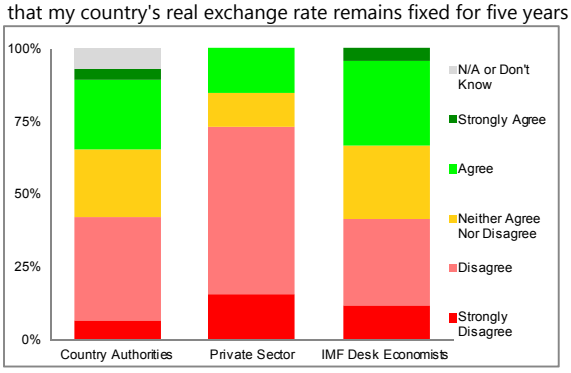
(d) Medium-term forecasts entail too much uncertainty to be valuable for policy discussions



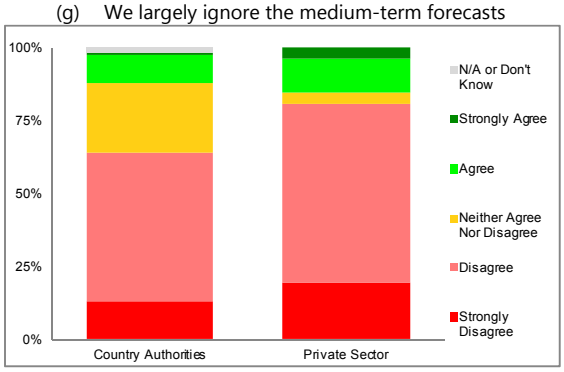
(e) When making medium-term forecasts, it is reasonable to assume that my country's output gap is closed by the end of five years



(f) When making medium-term forecasts, it is reasonable to assume that my country's real exchange rate remains fixed for five years



(g) We largely ignore the medium-term forecasts



Source: IEO Forecast evaluation survey.

Desk economists tend to value medium-term forecasts less than short-term forecasts for policy discussion purposes, and the opposite holds in the private sector. Country authorities seem neutral on the matter.

205. Asked to agree or disagree with the statement that "for policy discussions, medium-term forecasts are more valuable than one- or two-year forecasts," country authorities seem broadly divided (see prompt (b) in Figures 9 and 10). Roughly a third of them agrees or strongly agrees, while another third disagrees or strongly disagrees. Desk economists, on the other hand, clearly value medium-term forecasts less than short-term forecasts for policy discussions; relative to the authorities, a much smaller proportion of the staff economists agrees that medium-term are more valuable than short-term forecasts, and the lower average staff score is found to be statistically different from the authorities'. The finding that for policy discussions, authorities seem to appreciate medium-term forecasts, relative to short-term forecasts, more than do desk economists, is robust. It holds across different country groupings based on geographical location, country development stage, and recent IMF program participation (Annex 4). The differing perceptions of the two groups may indicate a potential disconnect between the "demand" for medium-term forecasts by the membership and its "supply" by IMF staff.

206. Perhaps not surprisingly, medium-term forecasts are viewed as more relevant by authorities from program countries and/or economies in lower development stages. The differences are generally statistically significant.[86] It is reasonable to think of these economies' need to assess their prospects of correcting the imbalances that lead to IMF programs or, more broadly, of speeding their development. Medium-term projections can provide key parameters for planning purposes and public finance administration in these economies, even if by global standards they are more volatile, making the forecasts more uncertain.

207. The private sector shows substantially more positive views on the relative usefulness of medium-term forecasts for policy discussions; more than half of respondents in this group actually rate the usefulness of medium-term forecasts higher than that of short-term forecasts (Figure 10).[87]

---

[86] When responses are viewed across IMF area departments, the authorities from countries in EUR clearly stand out, attaching the lowest relative value to medium-term forecasts. Among staff responses, on the other hand, there are almost no statistically differences across different partitions of the survey data (Annex 4). Scores are always lower than both the neutral level (i.e., = 3) and the scores from authorities, indicating a low relative appreciation for medium-term forecasts that is entrenched and widespread among IMF desk economists.

[87] This preference may help explain the superior performance of private sector forecasters relative to the IMF in medium-term forecasts of GDP growth, reported in Section V above.

Neither staff, the private sector, nor country authorities would recommend the IMF to place less emphasis on one- and two-year forecasts and more emphasis on medium-term forecasts.

208. Respondents were asked whether, "In its analysis, the IMF should place less emphasis on one and two-year forecasts and more emphasis on medium-term forecasts." Country authorities, the private sector, and especially staff do not favor shifting efforts from short- to medium-term forecasts at the IMF (see prompt (c) in Figures 9 and 10).

209. Again, the responses suggest that country authorities have a more favorable view of medium-term forecasts than do desk economists.[88] As in the case of prompt (b), this is particularly true for program countries, as well as emerging market economies and low-income countries. Indeed, we find a positive and statistically significant correlation between the authorities' responses to prompts (b) and (c): respondents who value medium-term relative to short-term forecasts are more likely to be in favor of the IMF de-emphasizing short-term relative to medium-term forecasts. No such correlation exists in desk economists' responses (Annex 4).

210. Among desk economists, the rate of agreement with this statement is notably lower than that among the authorities; the difference is statistically significant and robust to different partitions of the data by country program participation, level of development, and IMF area departments (Annex 4). When the evaluation team conducted follow-up interviews with 50 randomly selected IMF economists, among the 38 economists who provided an answer, 17 agreed that "the process of generating estimates of potential output and forecasts of GDP growth five years ahead [is] receiving the proper amount of attention at the Fund," while 13 disagreed, and 8 said they were not sure.

Both authorities and IMF desk economists are relatively neutral on whether medium-term forecasts are too uncertain to be valuable for policy discussions. The private sector respondents clearly disagree.

211. Desk economists and country authorities show very similar neutral responses to the statement that "Medium-term forecasts entail too much uncertainty to be valuable for policy discussions" (see prompt (d) in Figures 9 and 10).

212. The private sector, on the other hand, clearly values medium-term forecasts, despite the uncertainty typically involved. While roughly one-third of both the authorities and the desk economists disagree and one-third agree, about 70 percent of responses from the private sector disagree or strongly disagree with prompt (d).

[88] This result is also robust to different partitions of the data along program participation, level of development, and area departments (see Annex 4).

213. Disaggregating the authorities' responses by country group, we see a *higher* degree of concern about the uncertainty of medium-term forecasts among authorities from countries (i) currently under an IMF program, (ii) classified as emerging markets and low-income countries, and (iii) covered by the Fund's African Department. Interestingly, authorities from these same groups of economies also expressed *more positive* views on medium-term based on responses to (b) and (c), which may seem contradictory at first glance. Indeed, the aggregate data shows a negative correlation between the authorities answers to prompts (b)–(c) and their answers to prompt (d).[89]

214. This indicates that, *unconditionally*, concerns about the uncertainty of medium-term forecasts tend to be stronger among country officials who have more negative views of these forecasts. Only when the income level and the participation in IMF programs are taken into account, respondents who attach more value to medium-term forecasts according to answers to prompts (b)–(c)—countries in lower levels of development, in AFR, and program-countries—also tend to be more concerned about the uncertainty in these forecasts. In these cases it may be *because* authorities in these countries value medium-term forecasts relatively more, they are more concerned about their accuracy.

Desk economists, country authorities, and the private sector seem comfortable with the assumption that output gaps close within five years.

215. Both country authorities and staff, with no significant statistical difference between them, generally agree that "When making medium-term forecasts, it is reasonable to assume that my country's output gap is closed by the end of five years" (see prompt (e) in Figures 9 and 10). The private sector is also generally comfortable with this assumption.

216. The assumption is not uncontroversial, however. Recall that we found evidence that this rule may have been introducing optimistic biases in forecasts. About 40 percent of both authorities and staff either "agree" or "strongly agree," while one-quarter and one-third, respectively, "disagree" or strongly "disagree" with it.[90]

Staff, authorities, and especially representatives from the private sector do not think it is reasonable to assume the real exchange rate remains fixed for five years when making medium-term forecasts.

217. All groups of respondents disapprove of assuming that the real exchange rate (RER) remains fixed for five years when conducting medium-term forecasts (see prompt (f) in Figures 9 and 10). Among both IMF desk economists and country authorities, with no

[89] See Figure A4.2 and Table A4.6(b), in Annex 4.

[90] There is little evidence of statistical differences in the responses from authorities and desk economists in different area departments, although economists in AFR and in MCD seem, respectively, less and more likely to support the assumption.

statistically significant difference between them, more than 40 percent disapprove of the exchange-rate assumption, while 36 percent and 27 percent, respectively, either agree or strongly agree with it. The assumption is even more strongly rejected in the private sector, where about 73 percent of the respondents disagree or strongly disagree with it.

218. Just as for the assumptions on the global economy (survey prompt (a)) and the closure of the output gap (prompt (e)), we find no organizational bias in the responses to the assumption on exchange-rate behavior.[91]

Country authorities and the private sector clearly do not ignore medium-term forecasts.

219. Faced with the statement, "We largely ignore the medium-term forecasts" about two-thirds of the authorities and 80 percent of the private sector respondents disagree or strongly disagree, clearly suggesting that medium-term forecasts are not ignored by stakeholders (see prompt (g) in Figures 9 and 10).

220. Among the country authorities, those from program countries (Table A4.2, Annex 4) and from emerging markets and low-income countries (Table A4.3) are statistically less likely than authorities in non-program countries and advanced economies to ignore medium-term forecasts. Across IMF area departments, economies in Asia and the Pacific are the least supportive (more likely to ignore the forecasts), while those in the Middle East and Central Asia are the most supportive of medium-term forecasts (Tables A4.4 and A4.5 (a)-(b)).

221. The higher disregard for medium-term forecasts by authorities from Asian and Pacific countries may be related to the fact that the IMF's forecasts for that region are clearly less accurate than forecasts by the private sector, as discussed in Section V above.

B. Summary of Survey Findings

222. Findings on users' and staff perceptions of IMF medium-term forecasts can be summarized as follows:

- Among the respondents to the evaluation survey, only the representatives from the private sector revealed a clear *absolute* preference for medium- over short-term

[91] When responses are disaggregated, few statistically significant differences are found across country groupings. Desk economists in advanced economies most clearly disagree with the assumption of a five-year fixed RER (differences from emerging market economies and LICs are statistically significant) and authorities in MCD are the most comfortable with that assumption, perhaps because of the widespread practices of pegged exchange-rate regimes in the region. See Tables A4.3 and A4.4 in Annex 4.

forecasts for policy discussions.[92] Relative to short-term forecasts, authorities have more positive views on medium-term forecasts than IMF desk economists.

- Very few respondents in any group say that they ignore medium-term forecasts, but most do not think that the IMF should place more emphasis on these forecasts at the expense of shorter-term forecasts.

- About one-third of the survey respondents from country authorities agree that medium-term forecasts are more valuable for policy discussions than short-term forecasts. An equal percentage disagrees. In phone interviews, however, virtually all participant country authorities clearly stated that they pay more attention to IMF short-term forecasts.

- Compared to both desk economists and country authorities, the private sector has the most positive views about the importance of medium-term forecasts for policy discussions. Only a small percentage of respondents in this group view the uncertainty in medium-term forecasts as a major obstacle to their use in such discussions. This may explain the superior performance of private sector forecasters relative to the IMF in medium-term forecasts of GDP growth, reported in Section V.

- There is little evidence of an organizational bias regarding the assumptions—about the global economy, the closing of the output gap, and the real exchange rate—that underlie IMF medium-term forecasts. In this regard, the views of country authorities and IMF desk economists are largely aligned: a clear majority of the respondents who expressed a preference agree that the assumptions about the global economy and that the output gap closes after five years is reasonable, while a roughly similar percentage thinks the assumption of a constant real exchange rate is inappropriate in the context of medium-term forecasts.

- The private sector is even more clear-cut in its judgments; only 10 percent of respondents think the assumption of a constant real exchange rate is reasonable, and a majority agree with the idea of basing forecasts on a closing of the output gap at a five-year horizon.

VII. Concluding Remarks and Lessons

223. This section summarizes and discusses the implications of the main findings and offers recommendations for strengthening the forecasting process.

[92] According to country authorities interviewed by telephone, short-term forecasts are also the most visible in the press and, therefore, entail more immediate political interest, which may tilt the views of authorities and desk economists towards them.

A. Main Findings

Forecast methods

224. For generating medium-term forecasts there are no guidelines about which methods IMF economists should use. Relative to statistical and structural methods, "judgment" is by far the most frequent method of choice, sometimes in combination with other methods. Some of the methods, such as the Hodrick-Prescott (HP) filter, frequently used by IMF economists to construct estimates of potential output that guide medium-term forecasting, may suffer from fundamental shortcomings that imply large degrees of uncertainty about these estimates. This uncertainty is not always communicated to the users of IMF forecasts as it should.

225. The general approach, including the methods used in medium-term forecasts of GDP growth for individual economies, seems broadly appropriate, but sufficient coordination for better analytical consistency at the regional and global levels may be lacking. Relative to short-term forecasts there is more room for improvement in terms of enhancing the top-down aspect of the process of constructing medium-term forecasts and better incorporating the interconnections and spillovers that exist in the global economy to these forecasts.

226. Results from the IEO evaluation survey of desk economists, country authorities, and the private sector, as well as from follow-up interviews, show little evidence of an organizational bias regarding the assumptions that underlie IMF medium-term forecasts. All groups of respondents seem comfortable with the assumptions about the state of the global economy and that output gaps close within five years; they all seem uncomfortable with the assumption that the real exchange rate stays constant for five years.

Informational efficiency and accuracy

227. Focusing on point forecasts, the overall results from the statistical analysis indicate that IMF medium-term forecasts for GDP growth in the *WEO* meet the basic forecasting efficiency standards in most countries, with little evidence of a built-in organizational bias. This is consistent with the findings by Genberg and Martinez (2014b) for short-tem forecasts and broadly in line with those by Timmerman (2006).

228. There is a tendency to overpredict GDP growth, however. This optimistic bias can be seen across all IMF area departments—with the notable exceptions of the Middle East and Central Asia Department and oil exporters—as well as in countries with different levels of development and regardless of their IMF program status.

229. The optimistic biases are found in a non-negligible share of the membership, including most G20 economies: in a universe of about 180 countries, between two-third and three-fourth of them show predicted growth rates, on average, higher than actual growth rates. In about 20–30 percent of countries this optimistic bias is statistically significant, although small sample distortions in standard tests indicate that the frequency of bias may be lower than that.

230. The average optimistic bias ranges between 0.14 percentage points and 0.76 percentage points, depending on the forecast horizon, the measure used (median or mean), and the method. Many of the countries showing biases in medium-term forecasts have been previously reported by Timmerman (2006), mainly in the context of short-term forecasts. Statistically significant biases are more frequent among economies in Africa, as well as in emerging-market and low-income economies.

231. Biases are less frequent but seem to be larger in the advanced economies as a group, reflecting the fact that large biases are concentrated in a few G7 economies, as previously reported by Timmermann (2006). The average optimistic bias we find for G7 economies as a group is always substantially larger than the overall bias in the full sample. Biases in the emerging-market economies group, on the other hand, are always the smallest.

232. Considering only statistically significant results, medium-term forecasts of GDP growth are unbiased for most (70–80 percent) of IMF member countries. In addition, despite being found in all area departments, the overoptimistic biases do not seem *systemic*, caused by any intrinsic problem with the way the IMF produces forecasts. For instance, these biases are critically influenced by an entrenched inability to predict recessions, which is not peculiar to the IMF.

233. When controlling for three important recession years—1997 (Asian Crisis), 2001 (terrorist attack in the U.S. and burst of the "dot-com bubble"), and 2009 (recent financial crisis)—the measured optimistic biases are greatly affected, being either substantially reduced, completely eliminated, or reversed.

234. Serial correlation is not a common problem, even less frequent than bias, which limits the scope to use past forecast errors to improve IMF medium-term forecasts. On the other hand, for an important share of member countries, there are signs of information inefficiency caused by the failure to internalize, into forecasts for these countries, the forecasts of GDP growth in large economies (United States, Germany, or China) and of oil prices. This type of informational inefficiency in IMF forecasts—disregard of potential economic interdependencies and international spillovers in the global economy—is much more frequent in medium-term than in short-term forecasts, suggesting that efforts to account for those interconnections have not been implemented in the former as they have in the latter.

235. In terms of accuracy, IMF medium-term forecasts can be seen as up to standard. They are clearly superior to naïve forecasts based on mechanical methods (suggesting that staff judgment adds quality to forecasts) and are only marginally less accurate than private sector (Consensus) forecasts.

236. Evidence in this paper, as well as from IMF research, shows that incorporating information on (i) interconnections in the global economy and (ii) "catch-up" forces in per capita income among peer economies can potentially improve the general quality of forecasts. In this regard, and in relation to the benefits of a more centralized approach that

enhances the cross-country consistency of forecasts, important lessons can be learned from the approaches to medium-term forecasts followed in other multi-country institutions such as the European Commission and the OECD, while recognizing that the greater heterogeneity of IMF member economies necessitates the use of more country-specific methods.

IMF programs

237. Over optimism in medium-term forecasts of GDP growth is found regardless of program participation status at the time of the forecast. However, on the question of whether programs entail additional optimism relative to non-program situations, a more nuanced picture emerges relative to what is discussed in the literature.

238. For "program countries"—defined as countries that have participated in at least one IMF program during the sample period—the medium-term forecasts show a clear additional optimistic bias relative to those for countries with no history of programs in the sample period. This additional bias occurs regardless of whether program countries are in a program or not at the time of the forecast. They are sizeable, statistically significant, and not altered by changes in the forecast horizon, forecast vintage, or the effects of big recessions.

239. The nuance comes from the fact that, conditional on being classified as a program country, the optimistic biases in four- and five-year-ahead forecasts are *smaller*—and the differences are statistically significant—when the forecast is made *during* a program relative to when program-countries are *not* in a program at the time of the forecast. Only three-year-ahead forecasts show statistically larger optimistic biases during programs and, even then, only for the Fall vintages of the *WEO*.

240. In any case, if it exists, the additional optimism in forecasts during IMF programs mainly arises from the forecasts made in the year that programs start. The overoptimism in these initial-year forecasts, which is perhaps influenced by considerations that are more binding at program inception—and not directly related to the forecast method—is usually large, ranging between -0.4 percentage point and -1 percentage point. It is quickly eliminated after the first year of the program. The inability to predict the occurrence of recessions can, to some extent, explain the pattern of statistically significant biases around program dates.

Staff incentives and views about medium-term forecasts

241. The evaluation survey results suggest that IMF desk economists see medium-term forecasts of GDP growth as less important than short-term forecasts, in part because the former entail too much uncertainty for use in policy discussions. Although this view may not be that of the IMF as a whole, it raises the question whether desk economists need more encouragement to enhance their analysis of structural issues and use it to improve medium-term forecasts.

242. The private sector's more positive views on the value of medium-term forecasts for policy discussions may explain the somewhat superior performance of their medium-term forecasts of GDP growth relative to the IMF.

243. These findings about the attitude of desk economists towards medium-term forecasts should not be taken to imply that the IMF largely disregards longer-term analysis more broadly defined. Medium-term scenarios and risk analyses are increasingly important in IMF's flagship and bilateral surveillance reports. Since the year 2000, on average, *WEO* editions contain 1.3 sections that deal with medium/long-term subjects. In about two-third of these editions, and in 84 percent of a random sample of recent Article IV reports, there is at least one chapter dealing with the analysis of longer-term issues.

B. Discussion

244. The author leaves to IMF Management and the Board of Executive Directors the question of whether overoptimistic bias in medium-term forecasts for about 30 percent of the membership is large enough to require action. In this regard the caveats noted earlier (Section V) about the interpretation of the statistical findings should be kept in mind. It should also be noted that biases in forecasts are not necessarily the result of deficient forecasting practices *at the institutional level* relative to peers. In terms of accuracy, the IMF does not seem to perform significantly worse than its peers. Again, it is ultimately for IMF Management and the Board to decide whether the superior accuracy performance of Consensus over the IMF is significant enough to deserve action.

245. The perception from users of IMF forecasts, including medium-term forecasts of GDP growth, is overall positive.[93] Moreover, authorities generally trust the integrity of IMF forecasts and are satisfied both with the interaction with staff during forecast preparation in Article IV consultations and with the transparency of the forecasting process, which they judge to be free of political influence.

246. These considerations, however, do not reduce the importance of a careful analysis of the problems highlighted by this paper, with a view to reducing their occurrence whenever possible.

247. Medium-term scenarios and risk analyses are increasingly important in the IMF's flagship and bilateral surveillance reports. If the point forecasts for medium-term GDP growth—which ultimately drive the longer-term scenarios in several IMF products—are not given appropriate attention, the risk assessments associated with these analyses may be compromised. Problems in these central forecasts can have significant effects on the IMF analysis over longer time horizons, including in risk assessments. To the extent that these

[93] Genberg and Martinez (2014a) show that between 70 percent and 90 percent of country officials view the Fund's forecasts of GDP growth as "about right" rather than consistently too high or too low.

analyses are becoming increasingly important in several Fund products, problems in point forecasts will likely affect the policy advice to member countries.

248. Overall, the survey data and post-survey interviews indicate that IMF desk economists (though not necessarily the IMF as a whole) see medium-term forecasts as less relevant for policy discussions than do country authorities and the private sector. It seems that short-term forecasts are more carefully done, absorbing more resources, time, and effort, while medium-term (point) forecasts are, from the perspective of desk economists, almost a byproduct—sometimes only produced because they are integrated into the macro framework spreadsheet and required from desk economists during the *WEO* process, not because of their perceived importance by the forecaster.[94]

249. A general implication of the findings regarding the IMF medium-term forecasts—(i) problems in a non-negligible number of countries, (ii) slightly worse performance in comparison with private forecasters, (iii) relatively less interest from desk economists, vis-à-vis short-term forecasts, (iv) and low value attached to these forecasts by desk economists relative to other stakeholders—may be that more attention should be paid to medium-term forecasts, though not necessarily in exchange for less attention paid to short-term forecasts. In follow-up phone interviews with country officials, this general recommendation frequently came up, especially from authorities in low-income countries.

250. Based on the evaluation survey results, efforts to improve medium-term forecasts of GDP growth should be especially important for program countries and less developed economies. But because forecasts for advanced economies also tend to show large systematic optimistic biases, perhaps disproportionately affected by the "five-year-rule" used by IMF forecasters for the closing of the output gap, forecasts for these economies may also deserve closer attention.

C. Overall Assessment

251. In analyzing the results from the statistical analysis, one should keep in mind that, for a number of technical reasons—related to differences in sample, forecast horizon, how to treat data revisions when defining outturns, how to measure the biases, and the benchmarks used in comparisons—as well as because of the effect of learning, by the forecaster, about potentially changing economic structures—these results must always be taken with a healthy degree of skepticism.

[94] This finding is consistent with interview results from the IEO evaluation "The Role of the IMF as Trusted Advisor" (IEO, 2013b), whereby country authorities, particularly in emerging-market and low-income economies, highlighted that "the Fund paid too little attention to how to promote sustainable growth and employment."

252. Overall, the quality of IMF medium-term forecasts for GDP growth is acceptable, although there is more room for improvement relative to short-term forecasts. Central or point forecasts are appropriately complemented by a broader set of discussions about medium-term issues—including risk assessments described with fan charts and alternative scenarios to baseline projections based on the central forecasts. These discussions are increasingly important in IMF flagship documents and, like the point forecasts on which they are based, are highly valued by country authorities.

D. Recommendations

253. Given the importance of medium-term forecasts of GDP growth for critical products of the IMF (Section II), in light of the variety of methods potentially used at the IMF (Sections III and IV), and considering the results of the survey on the perceptions of users and staff about these forecasts (Section VI), we offer the following potential recommendations (with comments):

Recommendation 1: Consider increasing the consistency across the institution regarding the process of producing medium-term forecasts, perhaps with some flexible guidelines and the discussion of appropriate methods for groups of countries in similar situations regarding data availability and stage of development.

- While some degree of flexibility in the process of producing medium-term forecasts may be warranted, given the heterogeneity in the economic structures across the IMF membership, efforts to (i) increase methodological consistency of forecasts made for similar types of economies, and (ii) ensure economic consistency at the country, regional, and global levels, may prove beneficial.

- A higher degree of homogeneity in the production of forecasts could be desirable for products such as the Pilot EBA— that uses relative five-year-ahead forecasts of GDP growth—especially if less attention and resources are spent in medium-term relative to short-term forecasts, as seems to be the case according to survey evidence.

- It can also help in the learning and dissemination of best practices in forecasts for peer economies. Further economic and analytical consistency may also produce efficiency gains by fostering the use by IMF economists of information from other economies when forecasting their own country of assignment. This is so in light of the benefits to the quality of forecasts derived from incorporating both interconnections in the global economy and "catch up" effects stemming from economic forces associated with the convergence of per capita income among peer economies (see Recommendation 7, below).

Recommendation 2: Develop ready-to-use toolboxes, both to implement different methods and to evaluate their forecasting performance.

- By reducing the cost of changing methods, this could increase the incentives to do so, with a view to improve the performance of medium-term forecasts when required. Similarly to methods already used by desk economists, these toolboxes should not be applied mechanically, but be complemented by appropriate judgment. The recently released Staff Guidance Note for Public Debt Sustainability Analysis in Market-Access Countries (SM/13/86) provides some useful guidance on this direction and could be extended and complemented.

Recommendation 3: Create a cross-departmental unit to discuss IMF-wide views on medium-term forecasts and the evolution of the global and regional economies over longer horizons—relying on the structural drivers of economic growth in the major regions and countries—as well as to monitor the IMF's medium-term forecasts at the country and regional levels.

- This unit could be molded after the Interdepartmental Forecast Committee, which ensure such unified views in the context of short-term forecasts (see Genberg, Martinez, and Salemi, 2014), perhaps with less frequent meetings.

254. The results of the forecasting performance tests are consistent with findings of previous evaluations (Section V).[95] In light of this, we mainly reinforce some recommendations made by Timmermann (2006) for the forecast process at the IMF, in general, and add two that are specific for medium-term forecasts:

Recommendation 4: Selectively devote more resources to medium-term forecasts—although not by reducing the resources dedicated to short-term forecasts.

- This could contribute to closing the gap between the "demand" from users of these forecasts, especially in less developed and program economies, who more clearly see medium-term forecasts as a useful input into policy making, and the "supply" of forecasts by desk economists, who tend not to value such forecasts as highly.

Recommendation 5: Continuously monitor forecasting performance, perhaps through a cross-departmental committee for medium-term forecasts as proposed in Section IV above.

- This recommendation is likely to be important across the IMF membership: for countries covered by the Asia and Pacific Department (for which forecasts are clearly

[95] See Artis (1988, 1996), Timmermann (2006), Freedman (2014), and Genberg and Martinez (2014b).

less accurate than those from Consensus), the African Department, low-income economies (where biases are most frequent), and advanced economies (where biases are larger, on average).

Recommendation 6: Create a culture of learning from past forecast errors.

- Interviews with desk economists showed that past errors are rarely looked at. Bias-adjusted forecasts should be used as guidance, not to mechanically "correct" forecasts but to inform the learning process and address the systematic, and sometimes persistent, errors in forecasts found for some countries. In forecasts of economies undergoing structural change, such efforts to reduce bias should take into consideration that biases may be changed over time by new forecasting practices. Keeping this in mind may help to avoid over-correction of forecasts for which biases might be getting smaller over time, partially in light of recent enhanced process and methods in IMF forecasts.

Recommendation 7: Take steps to better incorporate the notion of conditional convergence of per capita income across similar countries.

- This can be a way to both (i) increase the consistency of approaches used by staff to conduct medium-term forecasts of GDP growth and (ii) improve the quality of these forecasts along the lines discussed in some in-house research that deserves more attention—see Batista and Zalduendo (2004) and IMF (2004).

REFERENCES

Adolfson, M., and others, 2007, "Modern Forecasting Models in Action: Improving Macroeconomic Analyses at Central Banks," *International Journal of Central Banking,* Vol. 3 (4), pp. 111-144.

Adolfson, M., Linde, J., and Villani, M., 2007, "Forecasting Performance of an Open Economy Dynamic Stochastics General Equilibrium," *Econometric Reviews,* Vol. 26 (2-4), pp. 289-328.

Andrle, M., 2013, "What is in Your Output Gap?" IMF Working Paper 13/105 (Washington: International Monetary Fund).

Artis, M. J., 1988, "How Accurate Is the World Economic Outlook? A Post Mortem on Short-Term Forecasting at the International Monetary Fund," Staff Studies for the World Economic Outlook, World Economic and Financial Surveys, pp. 1-49 (Washington: International Monetary Fund).

Artis, M. J., 1996, "How Accurate Are the IMF's Short-Term Forecasts? Another Examination of the World Economic Outlook," IMF Working Paper 96/89 (Washington: International Monetary Fund).

Batista, C., and Zalduendo, J., 2004, "Can the IMF's Medium-Term Growth Projections be Improved?" IMF Working Paper 04/203 (Washington: International Monetary Fund).

Baxter, M., 1994, "Real Exchange Rates and Real Interest Differentials: Have We Missed the Business-Cycle Relationship?" *Journal of Monetary Economics,* Vol. 33, pp. 5-37.

Baxter, M., and King, R. G., 1999, "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series," *The Review of Economics and Statistics,* Vol. 81 (4), pp. 575-593.

Benes, J., and others, 2010, "Estimating Potential Output with a Multivariate Filter," IMF Working Paper 10/285 (Washington: International Monetary Fund).

Beveridge, S., and Nelson, C. R., 1981, A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'. *Journal of Monetary Economics,* Vol. 7, pp. 151-174.

Blanchard, O. J., and Quah, D., 1989, "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review,* Vol. 79 (4), pp. 654-673.

Borio, C., Disyatat, P., and Juselius, M., 2013, "Rethinking Potential Output: Embedding Information about the Financial Cycle," *BIS Working Papers, 404.*

Canova, F., 1998, "Detrending and Business Cycle Facts," *Journal of Monetary Economics,* Vol. 41, pp. 475-512.

——————, 1994. Detrending and Turning Points. *European Economic Review,* Vol. 38, pp. 614-623.

Cerra, V., and S.C. Saxena, 2000, "Alternative Methods of Estimating Potential Output and the Output Gap: Application to Sweden," *IMF Working Papers, No. 00/59* (Washington: International Monetary Fund).

Chauvet, M., and S. Potter, Forthcoming, "Forecasting Output," In *Handbook of Economic Forecasting (Vol. 2).*

Cheng, K. C., R. Duval, and L. Zhang, 2013, "Trend Growth in Emerging Asia," *Staff Note (Unpublished).*

Christiano, L. J., and T.J. Fitzgerald, 2003, "The Band Pass Filter," *International Economic Review,* Vol. 44 (2), pp. 435-465.

Christiano, L. J., M. Eichenbaum, and C.L. Evans, 2005, "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy,* Vol. 113 (1), pp. 1-45.

Christoffel, K., G. Coenen, and A. Warne, 2010, "Forecasting with DSGE," *Working Paper Series (1185).*

Clarida, R. H., and M.P. Taylor, 2003, "Nonlinear Permanent–Temporary Decompositions in Macroeconomics and Finance," *The Economic Journal,* Vol. 113 (486), pp. C125-C139.

Clarida, R., and J. Gali, 1994, "Sources of Real Exchange-Rate Fluctuations: How Important are Nominal Shocks?" *Carnegie-Rochester Conference Series on Public Policy. 41, pp. 1-56.* North-Holland.

Clark, P. K., 1989, "Trend Reversion in Real Output and Unemployment," *Journal of Econometrics,* Vol. 40 (1), pp. 15-32.

Coenen, G., P. McAdam, and R. Straub, 2008, "Tax Reform and Labor-Market Performance in the Euro Area: A Simulation-Based Analysis Using the New Area-Wide Model," *Journal of Economic Dynamics and Control,* Vol. 32 (8), pp. 2543-2583.

Cogley, T., and J.M Nason, 1995, "Effects of the Hodrick-Prescott Filter on Trend and Difference Stationary Time Series: Implications for Business Cycle Research," *Journal of Economic Dynamics and Control,* Vol. 19, pp. 253-278.

Côté, D. and D. Hostland, 1996, "An Econometric Examination of the Trend Unemployment Rate in Canada," *Bank of Canada Working Paper 96-7.*

Cuddington, J. T., and A. Winters, 1987, "The Beveridge-Nelson Decomposition of Economic Time Series: A Quick Computational Method," *Journal of Monetary Economics,* Vol. 19, pp. 125-127.

D'Auria, F., and others, 2010, The Production Function Methodology for Calculating Potential Growth Rates and Output Gaps. *European Economy - Economic Papers (420).*

De Masi, P. R, 1997, "IMF Estimates of Potential Output: Theory and Practice," *IMF Working Paper, No. 97/177* (Washington: International Monetary Fund).

Del Negro, M., F. and Schorfheide, 2010, Bayesian Macro-econometrics, in *Handbook of Bayesian Econometrics*, ed. by H. K. van Dijk, G. Koop, and J. Geweke. Oxford University Press.

————— (Forthcoming), DSGE Model-Based Forecasting. In *Handbook of Economic Forecasting (Vol. 2).*

Edge, R., M. Kiley, and J.-P. Laforte, 2009, "Investment Shocks and the Business Cycle: The View from a Policy-Oriented DSGE Model," *Mimeo*, Federal Reserve Board.

Faust, J., 2013, "A Report of the Predictive Accuracy of the IMF's *WEO* Forecast," *Version as of February 5.*

Fellner, W. J., 1956, *Trends and Cycles in Economic Activity: An Introduction to Problems of Economic Growth.* New York: Holt.

Freedman, C., 2014, "An Evaluation of Commissioned Studies Assessing the Accuracy of IMF Forecasts," IEO Background Paper No. BP/14/02 (Washington: International Monetary Fund).

Genberg, H., and A. Martinez, 2014a, "User Perspectives on IMF Forecasts: Survey Methodology and Results," IEO Background Document 1 (Washington: International Monetary Fund).

Genberg, H., and A. Martinez, 2014b, "On the Accuracy and Efficiency of IMF Forecasts: A Survey and Some Extensions," IEO Background Paper No. BP/14/04 (Washington: International Monetary Fund).

Genberg, H., A. Martinez, and M. Salemi, 2014, "The IMF/*WEO* Forecast Process," IEO Background Paper No. BP/14/03 (Washington: International Monetary Fund).

Harvey, A. C., and A. Jaeger, 1993, "Detrending, Stylized Facts and the Business Cycle," *Journal of Applied Econometrics,* Vol. 8 (3), pp. 231-247.

Hamilton, J. D., 1994, *Time Series Analysis.* Princeton, New Jersey: Princeton University Press.

Hodrick, R. J., and E. C. Prescott, 1980, "Post-War U.S. Business Cycles: An Empirical Investigation," Working Paper (Pittsburgh: Carnegie-Mellon University).

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

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

Independent Evaluation Office of the International Monetary Fund, 2013b, "The Role of the IMF as Trusted Advisor" (Washington: International Monetary Fund).

International Monetary Fund, 2004, "Policy Formulation, Analytical Frameworks, and Program Design," Paper prepared by Staff of the Policy Development and Review Department.

International Monetary Fund, 2009, "Global Prospects and Policies," *World Economic Outlook (October).*

International Monetary Fund, 2013a, External Balance Assessment (EBA) Methodology: Technical Background. Paper prepared by Staff of the IMF's Research Department. Unpublished. [could not find this in text]

International Monetary Fund, 2013b, "Staff Guidance Note for Public Debt Sustainability Analysis in Market-Access Countries," SM/13/86 (Washington: International Monetary Fund).

Johansson, A., and others, 2013, "Long-Term Growth Scenarios," *OECD Economics Department Working Papers (1000).*

Johnson, C. A., 2013, "Potential Output and the Output Gap in Central America, Panama and Dominican Republic," *IMF Working Paper, No. 13/145* (Washington: International Monetary Fund).

Kalman, R. E., 1960, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering, Transactions of the ASME Series D, Vol. 82, 35-45.*

Kim, C.-J., and C.R. Nelson, 1999, *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications.* MIT Press Books.

King, R. G., C.I. Plosser, and S.T. Rebelo, 1988, "Production, Growth and Business Cycles: I. The Basic Neoclassical Model." *Journal of Monetary Economics,* Vol. 21 (2), pp. 195-232.

Kuttner, K. N., 1994, "Estimating Potential Output as a Latent Variable," *Journal of Business & Economic Statistics,* Vol. 12 (3), pp. 361-368.

Laxton, D., and R. Tetlow, 1992, "A Simple Multivariate Filter for the Measurement of Potential Output," *Technical Report,* Vol. 59.

Loungani, P., 2001, "How Accurate are Private Sector Forecasts? Cross-Country Evidence from Consensus Forecasts of Output Growth," *International Journal of Forecasting,* Vol. 17, pp. 419-432.

Lucas Jr, R. E., 1976, "Econometric Policy evaluation: A Critique," *Carnegie-Rochester Conference Series on Public Policy,* Vol. 1, pp. 19-46. North-Holland.

Luna, F., 2014, "IMF Forecasts in the Context of Program Countries," IEO Background Paper BP/14/05.

Matheson, T., 2013, "The Global Financial Crisis: An Anatomy of Global Growth," IMF Working Paper No. 13/76 (Washington: International Monetary Fund).

Miller, S. M., 1988, "The Beveridge-Nelson Decomposition of Economic Time Series: Another Economical Computational Method," *Journal of Monetary Economics,* Vol. 21, pp. 141-142.

Mise, E., T.H. Kim, and P. Newbold, 2003, "The Hodrick-Prescott Filter at Time Series Endpoints," *Discussion Paper in Economics, 03/08.*

Mitchell, W. C., 1927, Introductory pages to 'Business Cycles: The Problem and Its Setting' In NBER, *Business Cycles: The Problem and Its Setting (pp. 23-0).*

Mohr, M. 2005, "A Trend-Cycle (-Season) Filter," *ECB Working Paper Series (499).*

Morley, J., 2002, "A State–Space Approach to Calculating the Beveridge–Nelson Decomposition," *Economics Letters,* Vol. 75 (1), pp. 123-127.

Morley, J., and J. Piger, 2012, "The Asymmetric Business Cycle," *The Review of Economics and Statistics, 94 (1).*

Murchison, S., and A. Rennison, 2006, "ToTEM: The Bank of Canada's New Quarterly Projection Model," Bank of Canada.

Musso, A., and S. Phillips, 2002, "Comparing Projections and Outcomes of IMF-Supported Programs," *IMF Staff Papers,* Vol. 49 (1) (Washington: International Monetary Fund).

Nelson, C. R., and H. Kang, 1981, "Spurious Periodicity in Inappropriately Detrended Time Series," *Econometrica,* Vol. 49 (3), pp. 741-751.

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 V. Wieland, 2012, "Complexity and Monetary Policy," *International Macroeconomics CEPR Discussion Paper No. 9107.*

Pedersen, T. M., 2001, "The Hodrick–Prescott Filter, the Slutzky Effect, and the Distortionary Effect of Filters," *Journal of Economic Dynamics and Control,* Vol. 25 (8), pp. 1081-1011.

Ravn, M. O., and Uhlig, H., 2002, "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations," *The Review of Economics and Statistics,* Vol. 84 (2), pp. 371-380.

Rennison, A., 2003, "Comparing Alternative Output-Gap Estimators: A Monte Carlo Approach," *Bank of Canada Working Paper 2003-8.*

Smets, F., and R. Wouters, 2003, An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association,* Vol. 1 (5), pp. 1123-1175.

St-Amant, P., and S.V. Norden, 1997, Measurement of the Output Gap: *A Discussion of Recent Research at the Bank of Canada*. Bank of Canada.

Timmermann, A., 2006, "An Evaluation of the World Economic Outlook Forecasts," *IMF Working Paper* (Washington: International Monetary Fund).

Watson, M.W., 1986, "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics,* Vol. 2, pp. 147-174.

Wold, H., 1938, "A Study in the Analysis of Stationary Time Series" (Doctoral dissertation; Stockholm).

ANNEX 1. METHODS TO ESTIMATE POTENTIAL OUTPUT

This annex provides further details on some of the methods discussed in Section III. All the methods aim at decomposing the time-series of real GDP, y_t, into a trend, usually a smoothed component, y_t^*, which is interpreted as the level of potential output, and a residual series $\hat{y}_t = y_t - y_t^*$, the cyclical component, or output gap.[1] In the rest of this annex, we will use this notation.

A. Statistical Univariate Methods

The Hodrick-Prescott Filter (HP)

The Hodrick-Prescott (HP) Filter is a very simple and effective method, readily available in major statistical and econometric software, which may explain in part its widespread use as a standard de-trending method in macroeconomics despite both heavy criticism—see Harvey and Jaeger (1993), Canova (1994, 1998), Cogley and Nason (1995), St-Amant and Van Norden (1997), and Mise, Kim, and Newbold (2003)—and the subsequent development of more sophisticated band-pass filters—for example, Baxter (1994), Baxter and King (1999), and Christiano and Fitzgerald (1999).[2]

The HP filter is a two-sided linear smoothing method used to decompose any time-series y_t into a smoothed trend component, y_t^*, and a residual series $\hat{y}_t = y_t - y_t^*$, the cyclical component. When applied to an output series, say GDP, it is usual to treat y_t as the natural logarithm of GDP at time *t*, such that changes in (log of) output, $\Delta y_t = y_t - y_{t-1}$, correspond to the (approximate) growth rate of GDP, and the output gap, $\hat{y}_t$, is the (approximate) percentage deviation of the actual GDP from its trend.

---

[1] The first attempts to separate the trend from cycle in economic series (output in particular) may be traced to Mitchell (1927). Early methods—for example, Fellner (1956)—used deterministic functions of time, usually polynomials, to represent the trend, whereby the cycle component emerges as a residual from the trend line. Nelson and Kang (1981) discuss the pitfalls of these early analyses. Since then, the literature has substantially evolved, with many new, more sophisticated approaches being proposed, but no single method can be regarded as appropriate for all situations. Andrle (2013) discusses how several of the methods discussed in this Annex can be represented using a unified framework.

[2] Baxter and King (1999) provide a comparison of their band-pass filter (BPF) with other univariate filters. BPFs are designed to allow frequencies within a certain range and reject (fully or partially) frequencies outside that range. In the trend-cycle decomposition of GDP, BPFs rely on prior information about key characteristics of the cycle, such as its frequency and duration, to inform particular moving averaging techniques that will be consistent with these characteristics. These filters aim at minimizing the (square) differences relative to an "ideal" filter that completely removes, from the GDP series, all information that does not satisfy (i.e., it is outside the frequency bands defined by) the researcher's prior views about the cycle. Baxter and King (1999), for example, suggest removing all information outside the frequency band between 6 and 32 quarters. They also show that, under certain assumptions about the cycle, the HP and BPF can produce very similar results.

The procedure minimizes the variance of the output gap ($\hat{y}_t$) subject to a penalty for the second difference of the trend (y_t^*). More specifically, using observations for the raw series of y_t for periods $t = 0, 1,\ldots, T$, the HP filter solves for the value of y_t^* that minimizes:

$$\theta_t = \sum_{t=0}^{T}(y_t - y_t^*)^2 + \lambda \sum_{t=1}^{T-1}[(y_{t+1}^* - y_t^*) - (y_t^* - y_{t-1}^*)]^2 , \qquad \text{(A1.1)}$$

where the penalty parameter λ controls the smoothness of the trend component. The larger λ is, the smoother y_t^*. In the limit, as λ increases, y_t^* approaches a linear trend.[3]

However, due to some well-documented undesirable properties, a naïve, mechanical use of the HP filter to estimate potential output and the output gap may lead to distorted results. One known problem is the difficulty in properly calibrating the smoothing parameter λ that simultaneously controls the smoothness of the trend and the volatility of the output gap. The choice is ultimately arbitrary, despite some guidance offered in the literature. A value of λ that is too low may generate excessive cyclicality in the trend when there is none and a value that is too high may smooth out existing structural breaks. Because cycle and trend are determined simultaneously, conditional on the arbitrary value of λ, an excessively smooth trend necessarily leads to an excessive volatile cycle.

Another major drawback of the HP filter arises from its poor properties at the end of the sample period, including a high sensitivity to the addition of new data points. This end-sample bias originates from asymmetric weighting of information at the start and end of the sample. Only around mid-sample, with roughly an equal number of observations before and after, will the filter be a symmetric "two-sided" filter and behave closely to a centered moving average. Closer to the end of the sample, however, the weight placed on contemporaneous observations for determining the trend increases, and the filter becomes one-sided. At the same time, the trend for the observations next to the endpoints tends to attach a larger weight on the first or last observations than on themselves. These two factors combined are the origin of the endpoint sample problem: when the sample size gets revised at time *t*, and *k* new observations become available, the effect on the estimates for time *t* and periods at the end of the original sample (i.e., prior to the addition of new data) can be substantial. The bias occurs precisely where accuracy matters most for forecasting and policy decisions.

---

[3] A value typically used by researchers when using quarterly data is $\lambda = 1600$, following suggestions by Hodrick and Prescott (1980, 1997). This value is motivated as follows: the conditional expectation of y_t^*, given the observations of y_t, is the solution of the minimization of (A1.1) *if* the cyclical components and the second differences of the trend components are identically and independently distributed, normal variables with mean zero and variances σ_1^2 and σ_2^2 (which they are typically not), *and if* $\lambda = \sigma_1^2/\sigma_1^2$. Using their (arbitrary) "prior view that a 5 percent cyclical component is moderately large, as is a 1/8 of 1percent change in the growth rate of output in a quarter," they set $\lambda = \sigma_1^2/\sigma_1^2 = 5^2/(1/8)^2 = 1600$. Ravn and Uhlig (2002) provide further discussion for values used when data are available at annual frequency. Suggestions range from $\lambda = 6.5$ to $\lambda = 400$. Pedersen (2001) suggests a method to optimally identify λ.

Because of the high sensitivity to new observations at the end of sample, as more recent data become available, relying on the HP filter to estimate y_t^* and $\hat{y}_t$ may lead to the awkward situation where the estimated potential output and output gap for the past date *t-k*, computed at time *t*, may be very different from those computed at time *t-k*. New data points will change, sometimes dramatically, past measurements of y_t^* and $\hat{y}_t$. Even though this feature may be an unavoidable side-effect of the methodology, and interpreted as part of a natural "learning process" involved in the assessment of unobserved variables that should benefit from new information, such "revisions of the past" can be potentially large and completely change the views one may have had about the cyclical position of the economy then, which may have even motivated policy decisions that now, at time *t*, seem misguided. Section II of the main text shows that similar revisions of past estimates are common in IMF potential GDP and output gap estimates.

As an attempt to mitigate the HP filter's end-of-sample problems, researchers use different techniques to modify the weight structure of the filter at the end of sample. One popular method, recently used at the IMF by Johnson (2013), is to provide projections for future years, beyond the end of the sample, then HP-filter the data for the extended sample, pushing the end-of-sample bias to the extended part of the sample, which may be later discarded. Of course, the accuracy of this remedy depends crucially on the quality of the projections for the out-of-sample observations.

A common way of tentatively incorporating economic theory into modifications of the HP filter is by specifying a bivariate filter whereby inflation becomes related to the slackness in the economy through a standard Phillips Curve, measured by the HP filter's cyclical component. That equation's squared error is then introduced into the filter's objective function (A1.1). Laxton and Tetlow (1992), Côté and Hostland (1993), and Rennison (2003) apply this technique to Canadian data to extract measures of potential GDP linked to inflation and unemployment fluctuations.

Similar transformations of univariate into bivariate or multivariate filters by incorporating restrictions from economic theory are discussed below, in the context of State-Space (SS) models and multivariate filters. Borio, Disyatat, and Juselius (2013), for example, include financial variables that may help the identification of the cyclical state of the economy.

Other proposed modifications in the HP filter include treating trend and cycle completely separately. Mohr (2005) proposes a trend-cycle (TC) filter that modifies the HP filter by treating both the cyclical and trend components as separate components rather than treating the cyclical component as a residual that is simultaneously determined. In that case, the minimization problem in (A1.1) is altered to include an optimal path for the cycle and a third, residual component. By amending the HP filter with a model for the cycle, the TC filter replaces the smoothing parameter λ with an expression that relates to the cycle frequency and length. This modification also implies that the trend becomes cyclical-dependent, differently from the HP filter.

All these modifications are shown to reduce, but not completely eliminate, the end-of-sample bias and the frequent "revisions of the past" generated by the method.

The Beveridge-Nelson (BN) decomposition

Intuitively, the BN decomposition is based on the assumption that any forecast of the GDP level made at time *t* should no longer be influenced by its transitory component as the forecasting horizon approaches infinity, when the forecast should reflect only the trend component. An auxiliary forecasting model is needed to produce a series of forecasts, going from one period ahead to this very long-term horizon.

The BN decomposition de-trending method, like the HP filter, is "mechanical," in the sense of completely abstracting from economic theory when decomposing the observed series of output into the unobserved permanent (trend) and cyclical components. See Beveridge and Nelson (1981), Cuddington and Winters (1987), and Miller (1988). Unlike the HP filter, though, the trend component obtained with the BN method has two sub-components, a deterministic trend ($\bar{y}_t^*$) and a stochastic trend ($\tilde{y}_t^*$). The procedure assumes that the change in output, Δy_t, is stationary.

Since Δy_t is assumed stationary, Wold's theorem applies—see Wold (1938). It essentially means that the series of changes in output has an infinite-order moving-average (MA) representation and can be written as:

$$\Delta y_t = y_t - y_{t-1} = \mu + \sum_{i=0}^{\infty} \lambda_i \, \varepsilon_{t-i}, \quad \lambda_0 \equiv 1, \tag{A1.2}$$

where μ is the long-run average growth rate of GDP, $\varepsilon's$ are uncorrelated disturbances with mean zero and constant variance, and λ_i's are constants. The stationarity of Δy_t insures that the sum $\sum_{i=0}^{\infty} \lambda_i$ is convergent.

Iterating forward on (A1.2), it is possible to show that a *k*-step ahead forecast of y, using information up to time *t*, denoted $\Delta y_{t,t+k}^F$, can be represented as a linear function of the forecast horizon *k* with slope μ and an "intercept" that is stochastic, containing the current level of GDP and a function of all the past disturbances. Beveridge and Nelson (1981) note that when one considers very long forecast horizons, (i.e., *k* becomes very large), this term becomes independent from *k* and can be treated as a constant. In this limit situation, Beveridge and Nelson (1981) interpret this term as the permanent or trend component of y_t, and show that it evolves according to a random walk:

$$y_t^* - y_{t-1}^* = \mu + \left(\sum_{i=1}^{\infty} \lambda_i \right) \varepsilon_t, \;\; \lambda_0 \equiv 1, \tag{A1.3}$$

Note that, starting at $t = 0$ and iterating forward on (A1.3), gives:

$$y_t^* = y_0^* + \mu t + \Psi \sum_{j=1}^{t} \varepsilon_j \,, \quad \Psi \equiv \left(\sum_{i=1}^{\infty} \lambda_i \right), \tag{A1.4}$$

where $\bar{y}_t^* = y_0^* + \mu t$ is the deterministic trend and $\tilde{y}_t^* = \Psi \sum_{j=1}^{t} \varepsilon_j$ is the stochastic trend.

The cyclical component corresponds to the sum of all possible *j*-period ahead forecasts for the growth in output, $j = 1, 2, \ldots, \infty$, after excluding a term that accounts for the deterministic growth ($k\mu$):

$$\hat{y}_t = \lim_{k \to \infty} \sum_{j=1}^{k} \left[\Delta y_{t,t+j}^F - k\mu \right] \tag{A1.5}$$

where $\Delta y_{t,t+j}^F$ is the *j*-period ahead forecast of output growth.

That is, the BN trend is the long-horizon conditional forecast of the time series after discarding any deterministic drift effects. Any forecast made at time *t* should no longer be influenced by its transitory component as the forecasting horizon approaches infinity, when it should reflect only the trend component.

The practical implementation of the procedure, given the impossibility of computing the infinite sums in (A1.2), (A1.4), and (A1.5), involves two steps. First, the estimation of a rational representation of (A1.2), a mix of a *p*-th order autoregressive (AR) and a *q*-th order moving-average process—i.e., an ARMA(*p*,*q*) process:

$$\Delta y_t = \bar{\mu} + \sum_{i=1}^{p} \phi_i \, \Delta y_{t-i} + \sum_{j=0}^{q} \theta_j \, \varepsilon_{t-j} \,, \tag{A1.6}$$

where the truncation limits *p* and *q* are optimally selected using standard information criteria (Akaike, Schwarz, Hannan-Quinn, for example) and $\bar{\mu} = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_q)$. Equation (A1.6) relies on the stationarity of Δy_t, which assures that the infinite MA process (A1.2) can be inverted and has a rational form representation in the form of:

$$\Delta y_t = \mu + \frac{(1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q)}{(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)} \varepsilon_t \,, \tag{A1.7}$$

where the ϕ's and θ's are the AR and MA terms in (A1.6) and L is the lag operator such that, for any variable x_t, $L^k x_t = x_{t-k}$. See Hamilton (1994, pp. 26, 67-68).

Second, once estimates of parameters μ, ϕ_i's, and θ_j's are obtained, for each time t, a j-period ahead forecast, $\Delta y^F_{t,t+j}$, for j = 1,.., k, can be constructed using (A1.6). Beveridge and Nelson (1981) suggest forecasts up to k =100. These forecasts are then used in a practical version of (A1.5) to obtain the cyclical component. See also Morley (2002), Clarida and Taylor (2003), and Piger and Morley (2012).

One advantage of this method relative to the HP filter is that the BN decomposition is a one-sided procedure, only relying on past data to generate estimates of potential output for any point in the sample. That is, for a sample starting at time 0 and ending at time T, the BN decomposition uses data from time 0 to t-1 in order to compute potential output at time t < T, while the HP filter uses the whole sample, including future dates, which reduces its scope for forecasts in real time.

Although the BN method is also relatively simple to implement, accurate trend-cycle decomposition is conditional on the accuracy of the auxiliary forecasting model that is required to produce the forecasts for the very long horizon.

One disadvantage of this approach is that innovations to the cyclical and permanent components are perfectly correlated, which is hard to reconcile with economic theory and the widespread notion that the actual GDP is influenced by shocks—purely demand shocks, monetary policy shocks, for example—that may have little effect on potential output.

Kuttner (1994) points to an additional disadvantage of both the BN and HP methods: they typically provide no indication of the uncertainty around the estimates of potential output. Given the potentially large revisions in estimates of potential GDP growth, discussed in Section II, having a measure of the uncertainty could enhance the confidence in their use by policymakers.

Univariate state-space (SS) models

The common feature of univariate SS methods is the use of information about observed variables to estimate unobserved variables, by relying on a *state-space representation* of the underlying model. Applications of this method can be found in Watson (1986) and, more recently (at the IMF), Johnson (2013).[4]

[4] The latter applies this method to the estimation of potential output in Central America, Panama, and the Dominican Republic.

Watson (1986) proposes the following univariate model that, like the previous methods, does not take into account any considerations from economic theory:

$$y_t = y_t^* + \hat{y}_t \tag{A1.8}$$

$$\Delta y_t^* = \mu + \varepsilon_t \tag{A1.9}$$

$$\hat{y}_t = \phi_1 \hat{y}_{t-1} + \phi_2 \hat{y}_{t-2} + u_t \tag{A1.10}$$

As before, observed output, y_t, is the sum of trend (potential output, y_t^*) and cyclical (output gap, $\hat{y}_t$) components. Potential output is assumed to follow a random-walk process, while the output gap follows a second-order autoregressive, or AR(2), process. White-noise innovations ε_t and u_t represent permanent and transitory shocks to output, respectively, unlike the BN decomposition, where shocks ε_t and u_t are perfectly correlated.

The model can be written in a *state-space* form:

$$y_t = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_t^* \\ \hat{y}_t \\ \hat{y}_{t-1} \end{bmatrix} \tag{A1.11}$$

$$\begin{bmatrix} y_t^* \\ \hat{y}_t \\ \hat{y}_{t-1} \end{bmatrix} = \begin{bmatrix} \mu \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & \phi_1 & \phi_2 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1}^* \\ \hat{y}_{t-1} \\ \hat{y}_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ u_t \\ 0 \end{bmatrix} \tag{A1.12}$$

Equation (A1.11), which is equivalent to (A1.8), is called the *measurement equation*, and simply relates the observable (output) to the unobservable variables (potential output and output gap). System (A1.12), which consists of modified versions of equations (A1.9) and (A1.10), is the *transition* or *dynamic equation* that describes how the system evolves over time. Classes of models that have a state-space representation can be estimated by maximum likelihood, using the Kalman filter to evaluate the likelihood function.[5]

Models within this approach, also known as *latent variable* or *unobserved components* models, are flexible enough to include restrictions and behavioral equations taken from economic theory to help in the identification of potential output and the output gap, although this is not a requirement.

[5] The Kalman filter is a recursive method that uses prior information, starting with a guess, about the parameters to be estimated (in the case above, ϕ_1, ϕ_2, and μ plus the variances of ε_t and u_t) to construct (prior) estimates of all the variables for time *t* using information up to *t*-1. Once the information at time *t* is known (that is, there is a new measurement of the observable variable), then a posterior estimate of the variables for time *t* is constructed using the prior estimate plus a factor *k* (gain) of the prior measurement error, given by the observed value minus the prior estimate. The procedure is repeated until the entire sample is analyzed and a posterior estimate of all unobserved variables is obtained. See Kalman (1960) and Hamilton (1994, pp. 372-408).

For instance, Kuttner (1994) modifies Watson's univariate, a-theoretic version of this method by including an equation representing the Phillips curve, a systematic relation between the output gap and inflation (π), into the system above:

$$\Delta\pi_t = \mu_\pi + \gamma\Delta y_{t-1} + \beta\hat{y}_{t-1} + \nu_t \ , \qquad \text{(A1.13)}$$

where ν_t is a temporary disturbance specifically related to inflation.

Clark (1989) adds unemployment to the system of equations and assumes it has the same cyclical component as output, as a way of better identifying this component from the data. Similar modifications, that transform univariate methods in bivariate or multivariate approaches, are discussed in the next subsection.

In addition to being flexible to the inclusion of some economic theory in the estimation, there are at least two other advantages of univariate SS methods over the BN and HP filters. First, like the BN decomposition, but differently from the HP filter, the procedure can be used in real time as new data become available (potential output estimates at time *t* do not depend on future observations). Second, the method provides estimates of the uncertainty around the potential output series.

Univariate regime-switching (RS) models

Regime-switching models used in the estimation of potential output exploit the possibility that cyclical deviations from the trend, and the trend itself, may be generated by different stationary data generating processes. These different processes are associated with different regimes, or states, and the transition from one state to another is governed by a discrete set of probability distributions. Cyclical downturns and booms, for example, are considered two separate states and allowed to behave asymmetrically in terms of duration and depth.

In the vintage version of the method, these two separate regimes—downturns and booms—are assumed to be "unsustainable" in the sense that the unconditional probability of the economy staying in each of these regimes is much lower than in a third regime, which refers to the "sustainable" case, interpreted as underlying the evolution of potential output.

The method tries to uncover the probability distributions that govern the transition from one regime to another. Once the parameters of these underlying probability distributions are known (or estimated), one can draw on these distributions to obtain both the most likely path for the trend and confidence intervals. See Hamilton (1994) and Kim and Nelson (1999).

More concretely, assume that there are three possible regimes indexed by s_t. The basic assumption is that the realizations of output, y_t, are generated from three different probability distributions represented by density functions, associated, respectively, with periods when actual output is below (recessions, $s_t = 1$), at (sustainable path, $s_t = 2$), or above (booms, $s_t = 3$) its potential level:

$$f(y_t|s_t = j, Z_{t-1}; \Theta), \qquad j = 1,2,3, \tag{A1.14}$$

where Z_{t-1} represents all the information (including about s_t and y_t) available up to time *t*-1 and Θ represents the parameters of the probability distributions of both $y_t$and s_t.

The probability of being in one particular state, given Z_{t-1}, is represented by:

$$f(s_t = j|Z_{t-1}; \Theta) = \mu_j, \ j = 1,2,3, \tag{A1.15}$$

The density function (A1.15) is chosen such that the levels of GDP associated with $s_t = 1$ or $s_t = 3$ are unsustainable (that is, in the long run the probability of $s_t = 2$ dominates and the system tends to concentrate around this "absorbent" state). The joint probability of a particular realization of y_t and s_t can be obtained from the Bayes' Rule:

$$P(y_t, s_t = j|Z_{t-1}; \Theta) = \mu_j f(y_t|s_t = j, Z_{t-1}; \Theta), \ j = 1,2,3, \tag{A1.16}$$

Finally, one needs to integrate (A1.16) across all possible regimes in order to obtain the (marginal) probability distribution of y_t:

$$f(y_t|Z_{t-1}; \Theta) = \sum_{j=1}^{3} P(y_t, s_t = j|Z_{t-1}; \Theta) \ , \tag{A1.17}$$

When data on y_t are available for the sample $t = 1, \dots N$, estimates of the parameter set, Θ, can be obtained by selecting the value $\widehat{\Theta}$ that maximizes the following (log-) likelihood function:

$$\mathcal{L}(\Theta) = \sum_{t=1}^{N} f(y_t|Z_{t-1}; \Theta) \ , \tag{A1.18}$$

The estimated parameters can then be plugged into equations (A1.14)-(A1.17) to back out the underlying probability functions of s_t and y_t. By drawing on the probability distribution for the sustainable regime (i.e., $s_t = 2$) one can obtain both the most likely path for the trend and confidence intervals around it.

The pure switching regime approach has been recently used at the IMF by Johnson (2013). He finds that the method produces estimates of potential output growth similar to those from state-space models. The method is a purely statistical one, since no economic restrictions are required to help identify trend and cycle.

The clear advantage is that it can partially accommodate asymmetries in the frequency, depth, and length of recessions and booms, identified as periods in which GDP is below or above its trend, respectively. Morley and Piger (2012) use several specifications of this approach to

refine the shape (U-, V-, or L-shape) of recessions and compare the estimates of the trend based on these versions of the RS approach with those based on other approaches, such as simple autoregressive (AR) models, SS models, and the BN decomposition. Using information criteria, he finds that a regime-switching model in which the cycle does not show much variability in expansions but considerable variability in recessions is the preferred model. This finding implies an asymmetry in transitory fluctuations across expansion and recession phases, which are more likely to be caused by temporary and permanent shocks, respectively.

In principle, all the methods discussed in this subsection—HP, BN, univariate SS and RS models—can be used simultaneously and results can be averaged across models as a way to mitigate model uncertainty. For example, both Morley and Piger (2012) and Johnson (2013) report results from the averaging (Bayesian, in the former, simple average, in the latter) of the several procedures they use, both within RS methods and across different methods, as a way to reduce the effects of model uncertainty.

B. Multivariate Filters

To get an intuitive understanding of how univariate methods can be modified into bivariate or multivariate versions that incorporate insights from economic theory, consider a simple economy where inflation (π_t) is determined by the output gap together with other variables, some of which are not known by the analyst and are therefore captured by an error term u_t as in the following representation of the Phillips Curve:

$$\pi_t = \alpha\,(y_t - y_t^*) + u_t.$$

where the output gap is estimated as the difference between measured output (y_t) and the level of potential output (y_t^*) obtained using a univariate approach such as the HP filter, as in Laxton and Tetlow (1992), or the unobserved components SS model, as in Kuttner (1994).

If it were not for the error term it would be an easy matter to retrieve an estimate of potential output from the equation above if the sensitivity of inflation to the output gap (α) were known. But since α needs to be estimated, and since inflation is to some extent subject to unexplained shocks, additional information is needed. This is usually provided by an assumption about the behavior of potential GDP over time. A common assumption, but by no means the only possibility, is that potential output follows a simple autoregressive process as:

$$y_t^* = \rho y_{t-1}^* + \varepsilon_t.$$

From this example it is now easy to explain both the attractiveness and potential pitfalls associated with approaches that rely on some economic theory. The attractiveness is that they use economic relationships involving potential output to infer its value. But it is immediately obvious that this is an advantage only if the specified economic relationship provides a

reasonably good depiction of how the economy works; that is, if there are no serious model misspecification issues.

The equation specifying the evolution of potential output can also be a source of error. Is potential output well described by a regular process such as above or may it be subject to sudden jumps due to natural calamities, sudden technological innovations, or financial crises? The views the analyst has on these issues will have an impact on the final estimate.

However, there is no *a priori* reason to use only inflation as additional information to bring economic theory to the estimation of potential output. The bivariate model can be expanded to include other variables, such as unemployment, capacity utilization, and financial variables.

Benes and others (2010), in the IMF Research Department, propose a model that incorporates (i) information about the cyclical state of the economy extracted from the output gap, as well as reflected in the unemployment and capacity utilization rates, (2) identification restrictions based on the relationships between the output gap and these measures of economic slack—the Phillips Curve, in the case of inflation; Okun's Law, in the case of the unemployment rate; and a similar equation for capacity utilization—(3) long-term inflation expectations, and (4) long-run equilibrium relationships between potential output, the "natural" rate of unemployment, and capacity utilization.

One advantage of this method is its ability to exploit and ensure consistency between the estimated trend and several variables that are thought to be highly correlated with the business cycle. Moreover, the model can be estimated using Bayesian techniques, incorporating prior information about parameters and functional forms as a way to help in the identification of the parameters involved. Benes and others (2010) show that this method implies less revision of past estimates than the HP filter. Another advantage relative to some methods discussed above is that quantitative measures of the uncertainty around the estimates of potential output are possible.

The main drawbacks of this type of method, regarding its use for estimating potential output and medium-term forecasts, are its difficult implementation—substantially less straightforward than that of pure statistical filters—and the fact that the estimated long-run trends of several variables, including potential output, are highly dependent on the assumptions about the statistical properties of the trends of the variables in the model (more specifically, on how these trends evolve and how fast they converge to a stationary, usually constant, growth rate that is exogenously set). Depending on the assumptions, the path of potential output may sometimes imply an unrealistic speed of convergence of the estimated trend to the arbitrarily set steady-state path. Model misspecification and omitted variable biases are also sources of concern within this approach.

Models with financial variables

Potential output is defined by the concept of "sustainability," which is largely model-dependent. In a strict sense, because output cannot be persistently above the material limit given by the available resources in the economy; in an economic sense, because deviations from the sustainable level should trigger endogenous responses of prices (also wages, interest rates, exchange rates etc.) that would induce either a self-correction and/or the implementation of policies to help restore sustainability, at a pace that depends on the overall flexibility of these prices and the efficiency of the policies. The nexus between the output gap and inflation and the typical responses of central banks to inflationary pressures in situations of, say, overheating provides an illustration of how economists think about cyclical deviations from the potential, or sustainable, path of output. Borio, Disyatat, and Juselius (2013) argue that this reasoning may be incomplete.

The recent financial crisis raised the possibility that output might indeed have been on an unsustainable path prior to the crisis, even though inflation was low and stable. Borio, Disyatat, and Juselius (2013) speculate that the buildup of "financial imbalances" provided the fuel to maintain output above its potential level for a long time, only reverting towards potential output when these imbalances unwound in the wake of the financial crises. A different, but not mutually exclusive, view (IMF, 2009) argues that output may not return to its previous trend path following financial crises, but remain permanently below that trend. According to this alternative view, it is the potential output itself that falls after financial crises.

Regardless of which competing view represents the truth, both suggest that the estimation of the cyclical position of the economy relative to the trend should take financial variables into consideration. Ignoring the information content in financial variables about the cyclical position of the economy could produce wrong estimates of potential output. Borio, Disyatat, and Juselius (2013) show that by including financial variables in a multivariate filter framework, one can obtain more precise (in the sense of lower standard errors) estimates of potential output and output gaps that are also substantially more robust in real time and suffer less from the end-of-sample bias. They also suggest that the "finance-neutral" output gaps resulting from the proposed filtering technique produce more reliable estimates of cyclically adjusted fiscal balances and can be used more effectively as auxiliary guides for the conduct of monetary policy.

C. Structural Approaches

The main advantage of structural approaches is their clear economic interpretation of the evolution of the estimated output gap—driven by clearly defined structural shocks that have a zero mean in the long run. However, while the increase in detail, especially in the case of DSGE (dynamic stochastic general equilibrium) models, provides a more solid economic basis for the estimates, it also opens the door for model misspecification problems. As pointed out by Borio, Disyatat, and Juselius (2013), these methods' advantage in terms of

theoretical coherence is only meaningful to the extent that the underlying models are good approximations of reality. These methods are also much harder to implement.

DSGE models

DSGE models have become a very popular tool that policy makers, especially in central banks, use to evaluate alternative policy choices and track the effects of exogenous disturbances in different parts of the economy. They encompass a broad class of macroeconomic models that include the neoclassical growth model, basis for the Real Business Cycle (RBC) paradigm—see King, Plosser, and Rebelo (1988)—as well as models that account for many types of real and nominal rigidities, which are commonly known as New Keynesian models—see Christiano, Eichenbaum, and Evans (2005), and Smets and Wouters (2003).

The main characteristic of these models is their use of first principles, or explicit assumptions about how economic agents form their decisions and make their choices based on their preferences, subject to the available technology, and conditional on the institutional environment. The resulting framework "has a strong degree of theoretical coherence" (Del Negro and Schorfheide, forthcoming) that is less susceptible to the "Lucas Critique" than other forecast methods.[6] These models can therefore, in principle, be more effectively used in empirical analysis to assess the impact of policy making. Frequently estimated with Bayesian methods, DSGE models were first used for "story telling" and counterfactual exercises on the comparison of alternative policies, but are being increasingly used by central banks, with a mixed degree of success, in forecasting. Examples include the Bank of Canada's ToTeM (Murchison and Rennison, 2006), the Riksbank (Adolfson, Lindé, and Villani, 2007) and Adolfson and others, 2007), the New Area-Wide Model developed at the European Central Bank (Coenen and others, 2008; Christoffel and others, 2010), and the Federal Reserve Board (Edge and others, 2009). See Del Negro and Schorfheide (2010) for a review of the estimation of DSGE models.

Del Negro and Schorfheide (2012) show that, after some modifications to accommodate information about expectations of both inflation and the interest rate, as well as information from the Blue Chip Survey,[7] commonly used versions of DSGE models such as the Smets and Wouters' version have substantially better forecasting accuracy. They show that DSGE models still lose the accuracy contest against professional forecasts in the current- and next-

[6] The "critique" refers to Lucas (1976), which dismisses the predictions based on traditional macro-econometric forecasting models that rely on past correlations between macroeconomic variables because these correlations may change when new policies are introduced. In DSGE models, these correlations may also change, but the use of first principles assures that the changes can be ultimately traced down to "structural parameters" that govern preferences, technology, and the institutional environment.

[7] The Blue Chip Economic Indicators polls the forecasts made by several economists employed by the private sector about some macroeconomic indicators for the United States. Forecasts are provided for the current and next year. An average, or consensus, of their forecasts for each included variable is made available.

year forecasting horizons, but are competitive with, if not superior, to the Blue Chip consensus forecasts in the medium- and long-run forecasts for the United States.

In terms of the separation between trend and cycle, DSGE models work as a multivariate filter, whereby assumptions about the statistical properties of the trend are also required and will affect the estimates of potential output and output gap. Differently from that approach, however, the cyclical component of GDP is typically influenced by a larger set of different structural shocks (demand, supply, monetary, etc.) that are explicitly modeled, and by the much tighter restrictions imposed by economic theory.[8]

Structural VARs

Prior to the popularization of DSGE models, SVAR (Structural Vector Auto-Regressive) models were the reference among structural approaches.[9] Like DSGE models, this method also relies on a comprehensive set of equations that aim at representing the macroeconomy and uses several standard macroeconomic aggregates and variables in the estimation process. The equations usually included in the model aim at identifying (temporary) cyclical fluctuations with changes in aggregate demand or nominal shocks (such as monetary policy shocks) and the potential output with the (permanent) changes in aggregate supply. Examples are Blanchard and Quah (1989) and Clarida and Gali (1994).

The economy is represented by a system of reduced-form equations that express the relationship between the variables in the system and error terms that contain, but are not identical to, structural shocks—shocks that ultimately drive changes in all variables and have a more precise economic meaning. Identifying assumptions are needed to convert the reduced-form errors in the system into structural shocks. These assumptions are usually reflected in restrictions (economic theory) about contemporaneous correlations between variables and the sequence and timing in which different structural shocks affect each variable. Commonly used restrictions assume that output is affected by both supply and demand shocks in the short run, but only by supply shocks in the long run. Once the shocks are identified, potential output is computed from the permanent component of output in the system (i.e., the part only affected by permanent supply shocks).

Like DSGE models, SVARs do not suffer from end-of-sample bias but are susceptible to model misspecification issues where a problem in one equation in the model can be propagated to the whole system of equations. In the case of SVARs, model misspecification

[8] The equations that describe standard DSGE models are usually nonlinear. The most popular solution and estimation methods rely on a linear approximation of the system of equations around a particular point. This linear representation is usually expressed in the state-space form. In light of that, DSGE models can also be thought as multivariate SS models with more theoretical coherence and, likely, more details.

[9] The term "structural" has a stricter interpretation that only refers to economic models based on micro foundations, such as DSGE models, excluding SVARs.

can also emerge as a result of an inappropriate set of identification assumptions for the structural shocks. One obvious source of problems is the identification of potential output as only related to supply shocks, when demand shocks rooted in both private and public investment could also, in principle, have permanent effects on output.

Production function (PF) approach

Perhaps the analytical framework for introducing economic theory into the estimation of potential output in a more clearly and directly way is the production function (PF) approach. It postulates a specific relationship—i.e., a production function—between the output and the inputs, or production factors (labor, physical and human capital). The growth of output reflects either the accumulation of production factors or the increase in their productivity.

Since the quantity of production factors going into production in a given economy is, in principle, observable and measurable, the contribution to growth given by productivity is usually obtained by residual from the difference between the observed growth of output and the growth of inputs. This residual, sometimes referred as the "Solow residual," is a measure of total factor productivity (TFP).

More specifically, the standard production function approach postulates a given functional form, usually a simple two-factor Cobb-Douglas production function with constant returns to scale, to describe how production factors affect output (y_t): [10]

$$y_t = A_t K_t^{\alpha} L_t^{1-\alpha}, \tag{A1.19}$$

where A_t is the contribution of TFP to output; K_t and L_t are the stocks of capital and labor used in production, respectively; and α and $(1-\alpha)$ are the elasticities of output to changes in the stocks of capital and labor, respectively. The distinction between physical and human capital as well as the detailed decomposition of L_t, from working age population to effective hours worked (taking into account the labor participation rate, the unemployment rate, and the intensive margin of labor supply), are often used. See Cheng, Duval, and Zhang (2013) and D'Auria and others (2010), for example.

In its most basic version of the method, a postulated value for α, combined with direct measures or estimates of y_t, K_t, and L_t are plugged into (A1.19) to generate estimates of A_t as a residual. Then, estimates of long-run trends in A_t, K_t, and L_t are obtained, often using

[10] A theoretically motivated criticism of the PF approach relies on the implicit assumption of perfect competition in output markets that assures that the costs with production factors (i.e., income) completely span the output. That is, output should be equal to the income from labor and capital, and the parameter α should also represent the share of capital in total income, with $1-\alpha$ being the labor share. In reality, however, a portion of income is associated with rents derived from non-competitive markets and the calibration of α as the capital share may introduce an "omitted variable" bias.

one of the filtering techniques described in the previous subsections, although often K_t is not filtered. Denoting these trends with the superscript $(*)$, the potential output is then:

$$y_t^* = A_t(K_t^*)^{\alpha}(L_t^*)^{1-\alpha}. \quad \text{(A1.20)}$$

One important issue related to this method is that, in order to estimate the normal path, or trend, in output, estimates of the trend paths of both the inputs and TFP are required. These are then plugged into (A1.20) to generate the level of potential GDP.

However, if the univariate filtering techniques previously described are used to uncover the trend in production factors and productivity, as it is often the case, then the behavior of the estimated potential output using the PF approach will be very similar to—and, therefore, suffer from the same shortcomings as—that obtained from the direct filtering of the GDP series. That is, problems of GDP trend decomposition will shift to the trend estimates of the inputs. See, for example, Johnson (2013), who de-trends factors using the HP filter.

A second but related shortcoming of the simplest version of the PF approach is that interactions between cycle and trend, or short and long term, such as those implied by the Phillips Curve, as well as interactions between variables that may help the identification of the cyclical components (capacity utilization, unemployment rate, etc.), are not usually taken into account in the de-trending of the production factors and the TFP.

To address this issue, the comprehensive version of the PF approach described in D'Auria and others (2010), used by the European Commission (EC), relies on bivariate and multivariate filters to bring economic theory to bear in the identification of the trend in inputs and TFP. For instance, the EC uses a bivariate state-space model to exploit the links between the cyclical component of TFP and the degree of capacity utilization. This modification substantially reduces the likelihood of trend revisions in TFP as new data become available and generates smoother recoveries in trend TFP, in particular for the recent financial crisis, than estimates based on the HP filter. The authors argue that the more stable pattern for the trend TFP is more credible and reduces the degree of uncertainty embedded in the use of the PF approach for policy making. The EC's version of the PF approach also uses a model for the inflation-unemployment nexus that incorporates both price and wage rigidities, to help identify the trend in unemployment consistent with non-accelerating inflation—assumed to occur along the long-run path in which actual and potential output are identical.

Other common criticisms of the PF approach are the possible overly simplistic representation of the production technology and the fact that the estimates of potential output and the output gap are highly dependent on the estimates of the long-term, or natural, unemployment rate. Results are also sensitive to the "growth accounting" methodology used to generate the series of total factor productivity.

The vintage PF approach typically does not take into account spillovers from the global developments in technology progress and the distance to the technology frontier, nor does it

properly address slow moving demographic changes. Although this criticism can also be extended to all previously discussed methods, it is more pertinent in the case of the PF approach because of the roots of this approach are more directly linked to economic growth theory, where those factors take center stage.

ANNEX 2. STATISTICAL ANALYSIS OF IMF MEDIUM-TERM FORECASTS OF GDP GROWTH

This annex supplements the statistical assessment of the quality of IMF medium-term forecasts of GDP growth discussed in Section V. Details of the data, the different statistical tests, and the findings are provided.

A. The Data

The statistical analysis covers the period 1990–2012, using data from the *WEO* database. All IMF member countries for which there are more than six observations are included in the sample. In some cases, the sample is further trimmed to account for outliers and other potential problems with the data. These situations are explicitly identified.

For each year in the sample, the *WEO* publishes two vintages of forecasts, which are usually released at the IMF Spring Meetings (vintage 1) and, in the Fall, at the Annual Meetings (vintage 2). Let $e^{ij}_{t,t+k}$ be the error associated with a forecast, made at year *t*, vintage *j*, for the GDP growth rate at year *t*+*k*. The forecast error is computed as the difference between the actual realization (out-turn) of the GDP growth rate at year *t*+*k*, denoted g^i_{t+k}, and the *k*-year-ahead forecast made at year *t*, according to vintage *j*, $\hat{g}^{j}_{t,t+k}$. That is:

$$e^{ij}_{t,t+k} = g^i_{t+k} - \hat{g}^j_{t,t+k},$$

where superscript *i* on g^i_{t+k} indicates that, in order to accommodate data revisions, "actual" GDP growth at year *t*+*k* is only measured at some time in the future, at year *t*+*k*+*i*. This allows the data to be revised during *i* years before being treated as final and compared with predictions. Definitions of actual GDP growth for *i* = 1, 2 years are used. See also Faust (2013) and Genberg and Martinez (2014b).

Given the sample, the number of observations per country varies from a minimum of 6 to a maximum of 19, depending on the vintage (*j* = 1, 2), the definition of "actual" GDP growth (*i* =1, 2), and the forecast horizons (*k* = 3, 4, or 5 years). The first observations available on five-, four-, and three-year-ahead forecast errors (based on forecasts made in 1990) are computed for 1995, 1994, and 1993, respectively. The last observations are recorded for 2011 or 2010, depending on whether *i* =1 or 2, for forecasts made five, four, or three years before those years.

The time lag between forecasts and measures of actual realization implies that, with few exceptions, the sample does not include forecasts made during the recent financial crisis.

B. Tests for Unbiasedness

Two measures of the bias in forecasts are considered in the country-by-country descriptive analysis—the mean forecast error (μ) and the median forecast error (m).[1] Negative values of μ and m are indicative of optimistic bias—i.e., overprediction of GDP growth. Positive values indicate a pessimistic bias.

Standard (two-sided) t-tests are used to test the null hypothesis $H_0: \mu = 0$ and Wilcoxon tests are used to test $H_0: m = 0$. In both cases, the null hypotheses are tested against the alternative that the biases are different from zero.

One reason for the inclusion of a median-based measure of bias is that the median is a robust measure of the center of a distribution that may be skewed or have outliers. Another reason is that the Wilcoxon test for zero median does not require normality of errors.[2]

Following Timmermann (2006), regression-based tests of unbiasedness require the estimation of one of the following regressions for each country:

$$e^{ij}_{t,t+k} = \mu + \varepsilon_t \,, \tag{A2.1}$$

$$e^{ij}_{t,t+k} = \mu + \rho e^{ij}_{t-k,t} + \varepsilon_t \,. \tag{A2.2}$$

Since the regression residuals ε_t, have zero mean, an efficient forecast requires that $\mu = 0$ (unbiasedness) in (A2.1) and $\mu = \rho = 0$ (unbiasedness and absence of serial correlation) in (A2.2).

Tables A2.1 and A2.2 show summary descriptive statistics and regression-based statistics, respectively, for medium-term forecast errors in virtually the whole IMF membership (about180 countries).[3] The individual country statistics are available from the author upon request.[4]

---

[1] For a sample of size N, the mean forecast error is computed from:

$$\mu = \frac{1}{N}\sum_{t=1990}^{2012-i-k} e^{ij}_{t,t+k}, \quad k = 3,4,5, \quad i = 1,2 \,.$$

[2] In fact, the hypothesis of normal errors can be rejected (Jarque-Bera statistics) for virtually every country in the sample, regardless of the forecast vintage and the definition of actual GDP growth.

[3] Apart from the use of robust standard errors to test the hypothesis $H_0: \mu = 0$ in the regression-based tests, the test reported in Table A2.1 is essentially equivalent to that based on equation (A2.1) in Table A2.2.

[4] The frequency of statistically significant biases in Table A2.2, according to either regression (A2.1) or (A2.2), is slightly lower than in Table A2.1, due to the use of robust standard errors.

Note that the presence of outliers, combined with differences in the sample, partially explains why the cross-country average forecast errors seem to increase for shorter forecast horizons, going from a range of -0.5 percentage points to -0.6 percentage points, in five-year-ahead forecasts, to -0.6 percentage points to -0.8 percentage points in three-year-ahead forecasts (Table A2.1).[5] The sizeable differences in the average μ based on equation (A2.1) across different forecast horizons only hold when entire distributions are used; they virtually disappear when outliers and asymmetry are removed by truncating the distributions at one and two standard deviations (Table A2.2).

Table A2.1. Summary descriptive statistics
(Percent of countries, unless stated otherwise)

	5-Year-Ahead				4-Year-Ahead				3-Year-Ahead			
	1 Yr window		2 Yr window		1 Yr window		2 Yr window		1 Yr window		2 Yr window	
	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall
Number of countries	180	180	180	179	180	180	180	180	180	180	180	180
Average μ	-0.48	-0.56	-0.44	-0.45	-0.67	-0.67	-0.62	-0.53	-0.75	-0.76	-0.69	-0.64
$\mu < 0$	73%	73%	72%	72%	76%	76%	73%	73%	78%	78%	76%	74%
Rejects H_0: $\mu = 0$ (1)	32%	28%	31%	27%	31%	30%	28%	27%	33%	30%	26%	28%
Positive bias	25%	22%	21%	18%	26%	26%	21%	21%	31%	27%	23%	23%
Negative bias	7%	7%	10%	8%	4%	4%	7%	6%	2%	3%	2%	5%
Average m	-0.23	-0.24	-0.12	-0.14	-0.34	-0.33	-0.21	-0.18	-0.41	-0.37	-0.33	-0.24
$m < 0$	68%	67%	66%	64%	72%	67%	64%	65%	71%	67%	67%	62%
Rejects H_0: $m = 0$ (2)	33%	32%	31%	30%	33%	29%	29%	27%	32%	30%	23%	24%
Positive bias	27%	24%	22%	21%	28%	26%	21%	21%	30%	28%	21%	21%
Negative bias	6%	8%	9%	9%	5%	3%	8%	6%	2%	2%	2%	3%
Rejects H_0: $\rho = 0$ (1)	14%	18%	16%	22%	11%	11%	8%	8%	7%	7%	7%	9%

(1) Two-sided *t*-test; (2) Wilcoxon test.

Source: Author's calculations using *WEO*.

The country-by-country results reported in Tables A2.1 and A2.2 are obtained with fewer than 20 observations—sometimes as few as six. Given the small samples, in which the assumption normality is often violated, the accuracy and power of statistical tests may be questioned. Two strategies are used to account for small-sample distortions in the measurement of biases—bootstrap regressions and pooled least-squares estimations (PLS) of equation (A2.1).

First, following Timmermann (2006), the *p*-values from 1,000 bootstrapped versions of regressions (A2.1) and (A2.2) are computed for each country. Table A2.2 reports the

[5] When controlling for the same sample, this difference is slightly reduced.

percentage of countries for which the hypothesis of unbiasedness is rejected at the 5 percent significance level in more than 75 percent of the bootstrap regressions. For example, it shows that in 11 percent of countries, unbiasedness in five-year-ahead forecasts can be rejected more than 95 percent of the time and that in (11+6 =) 17 percent of countries, unbiasedness can be rejected more than 90 percent of the time. The information from bootstrapped *p*-values suggests that systematic biases may be even less widespread than the 20–30 percent of countries indicated by the standard tests.

Table A2.2. Summary Regression-Based Statistics
(Share of countries)

	5-Year-Ahead				4-Year-Ahead				3-Year-Ahead			
	1 Yr window		2 Yr window		1 Yr window		2 Yr window		1 Yr window		2 Yr window	
	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall
Equation (A2.1) # Obs	180	180	180	179	180	180	180	180	180	180	180	180
Average μ												
full distribution	-0.48	-0.56	-0.44	-0.45	-0.66	-0.67	-0.62	-0.53	-0.75	-0.76	-0.69	-0.64
within 2 std	-0.65	-0.68	-0.62	-0.61	-0.67	-0.71	-0.65	-0.67	-0.75	-0.73	-0.70	-0.71
within 1 std	-0.54	-0.53	-0.50	-0.51	-0.54	-0.50	-0.49	-0.50	-0.50	-0.49	-0.48	-0.47
Rejects H_0: $\mu = 0$	0.27	0.24	0.26	0.25	0.26	0.22	0.24	0.23	0.28	0.25	0.22	0.25
Bootstrap *p*-value < 0.05:												
[75%, 90%]	0.09	0.09	0.06	0.07	0.12	0.09	0.08	0.07	0.13	0.10	0.06	0.10
[90%, 95%]	0.06	0.02	0.04	0.04	0.02	0.03	0.04	0.03	0.03	0.05	0.03	0.05
[95%, 100%]	0.11	0.11	0.11	0.11	0.09	0.09	0.09	0.09	0.10	0.08	0.09	0.08
Equation (A2.2) # Obs	175	174	166	166	176	175	175	174	176	175	176	175
Rejects H_0: $\mu = 0$	0.33	0.29	0.31	0.28	0.28	0.27	0.23	0.25	0.20	0.21	0.19	0.21
Bootstrap *p*-value < 0.05:												
[75%, 90%]	0.14	0.10	0.10	0.08	0.10	0.07	0.09	0.06	0.09	0.05	0.07	0.05
[90%, 95%]	0.02	0.01	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02
[95%, 100%]	0.06	0.07	0.08	0.08	0.04	0.05	0.04	0.05	0.05	0.03	0.03	0.03
Rejects H_0: $\rho = 0$	0.16	0.18	0.19	0.20	0.13	0.10	0.10	0.13	0.09	0.11	0.10	0.11
Bootstrap *p*-value < 0.05:												
[75%, 90%]	0.03	0.02	0.02	0.02	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.01
[90%, 95%]	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
[95%, 100%]	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.02	0.00	0.00	0.00	0.00
Rejects H_0: $\mu = \rho = 0$	0.39	0.36	0.43	0.38	0.30	0.33	0.31	0.33	0.25	0.22	0.19	0.22
Bootstrap *p*-value < 0.05:												
[75%, 90%]	0.16	0.14	0.14	0.14	0.09	0.09	0.10	0.07	0.09	0.08	0.07	0.08
[90%, 95%]	0.03	0.03	0.05	0.05	0.03	0.03	0.05	0.03	0.02	0.03	0.03	0.03
[95%, 100%]	0.08	0.09	0.09	0.10	0.08	0.09	0.06	0.10	0.07	0.06	0.06	0.06

Note: Statistics based on two-sided *t*-tests.

Source: Author's calculations using *WEO*.

An alternative strategy to mitigate distortions due to small sample size is to use pooled country data in the regression-based analysis. Although insights for individual countries are lost, the PLS estimation of equation (A2.1) allows the use of fixed effects that take into account both country- and time-specific unobserved factors that may be important for the statistical tests.[6] More specifically, the following PLS regression is estimated:

$$e^{ij}_{t,t+k}(h) = \mu + \varepsilon_t(h) \tag{A2.3}$$

where h indexes the 180 countries potentially used in the sample and μ is the unconditional mean forecast error across countries and time, i.e., the measure of the bias. The timing convention used is that all forecasts are made at year *t*.

To estimate equation (A2.3) we proceed as follows (rather than excluding from the sample all observations from countries whose characteristics—such as large and frequent known structural breaks, poor data, etc.—would probably unfairly bias the results).

First, observations from transition economies of the former Soviet bloc in Eastern Europe and Central Asia, between 1990 and 1994, were excluded from the sample, because they were associated with large forecast errors that could be primarily associated with structural changes. Observations from these countries from 1995–2012 were kept in this first round. Observations from Afghanistan, Equatorial Guinea, and post–2003 Iraq were also excluded.

Second, outliers were identified using influence statistics. These statistics are used to measure the difference that any single observation makes to the estimation results of country-specific regressions based on equation (A2.1). Observations associated with large enough (absolute) values of the influential statistics were dropped from the sample as they are more likely to be outliers.[7]

---

[6] The use of fixed effects would substantially reduce much of the advantage of using the cross-country information if the regressions aimed at identifying *determinants* of the bias, because the statistical power to detect relationships between variables would be reduced. This could be an issue in the regressions that include dummy variables for the "big recessions" with global implications, discussed below. However, the results with and without the fixed effects are very similar.

[7] We consider the following influence statistic:

$$\bar{\varepsilon}_t(h) = \frac{\varepsilon_t(h)}{s_t(h)\sqrt{1-q_t(h)}} ,$$

where $\varepsilon_t(h)$ is the residual for observation at time *t* for country *h*, $s_t(h)$ is the variance of residuals that would have resulted had this single observation not been included in the estimation sample, and $q_t(h)$ is the *t*-th element of the partition related to country *h* in the "Hat Matrix," which is the projection matrix that contains information about the least-squares fitting. This statistic reflects the influence that any data point will have on the fitted values. Observations with $\bar{\varepsilon}_t(h) > 2$ were excluded.

Table A2.3 shows the results of the PLS regressions using the full sample (adjusted for outliers) and data from both *WEO* vintages (Spring and Fall), with out-turns measured after a two-year window. The estimated value of μ, which is the measure of bias, is reported in columns 2 and 7 for the Spring and Fall vintages, respectively, along with the (White cross-section-robust) p-values for the hypothesis $H_0: \mu = 0$ (columns 3-4 and 8-9). Two cases are considered: without fixed effects (columns 3 and 8) and with fixed effects (columns 4 and 9), both period and country-specific.

To investigate the effect of three "big recessions" with potential implications for a large share of the membership—the 1997 Asian Crisis, the 2001 recession associated with the terrorist attacks and the burst of the "dot-com" bubble in the United States, and the 2009 financial crisis—versions of equation (A2.3) are estimated with dummy variables that indicate when forecasts for 1997, 2001, and 2009 are made. For example, in the first row of Table A2.3, the dummy variable relative to the 2009 recession refers to (i.e., takes the value of one for) forecasts made in 2004 (five-year ahead), while in the second row it refers to forecasts made in 2005 (four-year ahead). The results for the estimated bias and associated p-values are shown in columns 5–6 (Spring vintages) and 10–11 (Fall vintages). [8]

Table A2.3. Pooled least squares estimation: average bias in GDP growth forecasts

	Spring, 2-year window					Fall, 2-year window				
		p-value		"Big" Recessions			*p*-value		"Big" Recessions	
Horizon	bias	Pooled LS	FE	bias	p-value	bias	Pooled LS	FE	bias	p-value
5-year-ahead	-0.33	0.0000	0.0000	-0.03	0.6132	-0.34	0.0000	0.0000	-0.03	0.6404
4-year-ahead	-0.44	0.0000	0.0000	-0.17	0.0045	-0.39	0.0000	0.0000	-0.12	0.0471
3-year-ahead	-0.53	0.0000	0.0000	-0.28	0.0000	-0.48	0.0000	0.0000	-0.22	0.0001

Source: Author's calculations using *WEO*.

Biases across country groupings

Tables A2.4, A2.5, and A2.6 show the full set of results for the average and frequency of statistically significant biases (at the 5 percent significance level) in five-, four-, and three-year-ahead forecasts of GDP growth according to four different measures within the country-by-country approach. Table A2.7 shows the results from the pooled least square estimations.

[8] When estimating a model with the dummy variables for the "big recessions," no time-fixed effects are included, only country-fixed effects.

Table A2.4. Frequency of statistically significant bias in 5-Year-ahead forecasts across different country groupings using the country-by-country approach
(Share of IMF member countries)

Descriptive Statistics		Mean				Median			
		1 Yr window		2 Yr window		1 Yr window		2 Yr window	
		Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall
Department	Full sample	0.32	0.28	0.31	0.27	0.33	0.32	0.31	0.30
	APD	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	AFR	0.08	0.08	0.09	0.07	0.09	0.10	0.09	0.09
	EUR	0.07	0.07	0.06	0.04	0.08	0.07	0.06	0.06
	MCD	0.03	0.03	0.05	0.06	0.03	0.04	0.05	0.05
	WHD	0.09	0.06	0.06	0.05	0.09	0.07	0.06	0.06
Development	AE	0.06	0.04	0.04	0.03	0.06	0.06	0.04	0.04
	EME	0.14	0.12	0.13	0.11	0.14	0.13	0.13	0.12
	LIC	0.12	0.12	0.14	0.13	0.12	0.13	0.14	0.14
	OECD	0.06	0.05	0.04	0.04	0.06	0.05	0.04	0.04
	G7	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.02
	G20	0.04	0.04	0.04	0.04	0.05	0.05	0.04	0.04
	OIL	0.02	0.02	0.03	0.03	0.02	0.02	0.03	0.03
	Primary	0.04	0.03	0.05	0.04	0.04	0.05	0.06	0.05
Regression-based statistics		Mean, equation (A2.1)				Mean, equation (A2.2)			
		1 Yr window		2 Yr window		1 Yr window		2 Yr window	
		Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall
Area department	Full sample	0.27	0.24	0.26	0.25	0.33	0.29	0.31	0.28
	APD	0.03	0.04	0.04	0.06	0.06	0.05	0.06	0.05
	AFR	0.08	0.07	0.07	0.06	0.08	0.07	0.09	0.07
	EUR	0.06	0.06	0.04	0.04	0.05	0.05	0.04	0.04
	MCD	0.03	0.03	0.04	0.04	0.06	0.05	0.06	0.05
	WHD	0.07	0.04	0.06	0.04	0.09	0.06	0.06	0.06
Development	AE	0.05	0.06	0.04	0.04	0.05	0.05	0.04	0.04
	EME	0.11	0.09	0.10	0.11	0.14	0.14	0.12	0.11
	LIC	0.11	0.10	0.12	0.10	0.15	0.10	0.16	0.13
	OECD	0.05	0.05	0.04	0.04	0.05	0.06	0.05	0.05
	G7	0.02	0.02	0.02	0.02	0.01	0.02	0.01	0.01
	G20	0.04	0.04	0.04	0.04	0.05	0.05	0.03	0.04
	OIL	0.02	0.02	0.02	0.04	0.03	0.03	0.04	0.03
	Primary	0.03	0.03	0.05	0.03	0.05	0.03	0.06	0.05

Note: The tests of unbiasedness rely on two-sided *t*-tests (mean) and Wilcoxon tests (median).

Source: Author's calculations using *WEO*.

Table A2.5. Frequency of statistically significant bias in 4-year-ahead forecasts across different country groupings using the country-by-country approach
(Share of IMF member countries)

Descriptive Statistics		Mean				Median			
		1 Yr window		2 Yr window		1 Yr window		2 Yr window	
		Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall
	Full sample	0.31	0.30	0.28	0.27	0.33	0.29	0.29	0.27
Department	APD	0.04	0.04	0.06	0.05	0.04	0.04	0.04	0.05
	AFR	0.07	0.07	0.07	0.07	0.09	0.08	0.08	0.08
	EUR	0.08	0.08	0.07	0.05	0.08	0.08	0.07	0.06
	MCD	0.03	0.04	0.04	0.04	0.03	0.03	0.04	0.04
	WHD	0.08	0.07	0.05	0.06	0.08	0.07	0.06	0.05
Development	AE	0.06	0.06	0.05	0.03	0.07	0.06	0.05	0.04
	EME	0.13	0.11	0.12	0.12	0.13	0.11	0.12	0.11
	LIC	0.12	0.13	0.11	0.12	0.13	0.12	0.12	0.12
	OECD	0.06	0.06	0.05	0.04	0.06	0.05	0.04	0.04
	G7	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.02
	G20	0.04	0.03	0.04	0.04	0.05	0.04	0.05	0.04
	OIL	0.02	0.03	0.03	0.03	0.02	0.03	0.02	0.03
	Primary	0.04	0.03	0.03	0.03	0.04	0.03	0.04	0.03
Regression-based statistics		Mean, equation (A2.1)				Mean, equation (A2.2)			
		1 Yr window		2 Yr window		1 Yr window		2 Yr window	
		Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall
	Full sample	0.26	0.22	0.24	0.23	0.28	0.27	0.23	0.25
Department	APD	0.04	0.04	0.04	0.04	0.03	0.04	0.02	0.03
	AFR	0.06	0.05	0.07	0.06	0.08	0.07	0.06	0.05
	EUR	0.07	0.06	0.04	0.04	0.05	0.05	0.03	0.03
	MCD	0.03	0.03	0.03	0.04	0.04	0.05	0.05	0.07
	WHD	0.07	0.05	0.06	0.04	0.09	0.07	0.07	0.06
Development	AE	0.06	0.06	0.04	0.03	0.04	0.04	0.03	0.03
	EME	0.11	0.08	0.08	0.09	0.11	0.11	0.10	0.11
	LIC	0.09	0.08	0.12	0.10	0.13	0.12	0.10	0.11
	OECD	0.06	0.05	0.03	0.03	0.05	0.05	0.04	0.03
	G7	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01
	G20	0.03	0.03	0.03	0.04	0.03	0.04	0.02	0.02
	OIL	0.02	0.02	0.02	0.03	0.01	0.02	0.03	0.05
	Primary	0.02	0.02	0.03	0.03	0.04	0.03	0.03	0.03

Note: Statistics based on two-sided *t*-tests (mean) and Wilcoxon tests (median).

Source: Author's calculations using *WEO*.

Table A2.6. Frequency of statistically significant bias in 3-year-ahead forecasts across different country groupings using the country-by-country approach
(Share of IMF member countries)

Descriptive Statistics		Mean				Median			
		1 Yr window		2 Yr window		1 Yr window		2 Yr window	
		Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall
	Full sample	0.33	0.30	0.26	0.28	0.32	0.30	0.23	0.24
Department	APD	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.04
	AFR	0.09	0.09	0.07	0.08	0.08	0.09	0.07	0.07
	EUR	0.09	0.08	0.08	0.07	0.11	0.08	0.07	0.05
	MCD	0.03	0.03	0.02	0.04	0.03	0.02	0.01	0.03
	WHD	0.08	0.07	0.05	0.06	0.07	0.07	0.05	0.05
Development	AE	0.07	0.06	0.06	0.04	0.08	0.07	0.06	0.04
	EME	0.13	0.10	0.08	0.11	0.12	0.09	0.07	0.08
	LIC	0.13	0.14	0.12	0.13	0.12	0.13	0.11	0.11
	OECD	0.06	0.06	0.06	0.06	0.07	0.06	0.06	0.05
	G7	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.02
	G20	0.04	0.03	0.03	0.04	0.05	0.04	0.04	0.04
	OIL	0.02	0.02	0.01	0.03	0.02	0.01	0.01	0.02
	Primary	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Regression-based statistics		Mean, equation (A2.1)				Mean, equation (A2.2)			
		1 Yr window		2 Yr window		1 Yr window		2 Yr window	
		Spring	Fall	Spring	Fall	Spring	Fall	Spring	Fall
	Full sample	0.28	0.25	0.22	0.21	0.20	0.21	0.19	0.18
Department	APD	0.03	0.03	0.03	0.03	0.00	0.01	0.01	0.02
	AFR	0.08	0.08	0.06	0.07	0.05	0.06	0.05	0.04
	EUR	0.08	0.07	0.06	0.04	0.05	0.06	0.06	0.06
	MCD	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.03
	WHD	0.06	0.05	0.04	0.04	0.07	0.05	0.05	0.04
Development	AE	0.07	0.06	0.05	0.04	0.03	0.04	0.04	0.04
	EME	0.11	0.07	0.07	0.07	0.07	0.08	0.05	0.06
	LIC	0.11	0.12	0.10	0.10	0.09	0.09	0.10	0.08
	OECD	0.06	0.06	0.05	0.04	0.04	0.05	0.05	0.04
	G7	0.02	0.02	0.02	0.02	0.01	0.01	0.02	0.02
	G20	0.04	0.04	0.04	0.03	0.02	0.02	0.02	0.02
	OIL	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.02
	Primary	0.03	0.03	0.04	0.03	0.02	0.02	0.02	0.01

Note: Statistics based on two-sided *t*-tests (mean) and Wilcoxon tests (median).

Source: Author's calculations using *WEO*.

Table A2.7. PLS Estimation: Bias in GDP Growth Forecasts across Area Departments and Development Stage

	Spring, 2 year window					Fall, 2 year window				
			p-value	"Big" Recessions				*p*-value	"Big" Recessions	
	Bias	Pooled LS	FE	bias	*p*-value	Bias	Pooled LS	FE	bias	*p*-value
					5-Year-Ahead					
AE	-0.66	0.0000	0.0000	-0.20	0.0406	-0.64	0.0000	0.0000	-0.18	0.0720
EME	-0.18	0.0782	0.0374	0.17	0.0808	-0.17	0.1113	0.0640	0.19	0.0632
LIC	-0.32	0.0017	0.0004	-0.15	0.1363	-0.36	0.0006	0.0001	-0.18	0.0723
AFR	-0.43	0.0004	0.0001	-0.32	0.0100	-0.46	0.0003	0.0001	-0.35	0.0054
APD	-0.33	0.0213	0.0047	0.10	0.4474	-0.30	0.0389	0.0103	0.12	0.3323
EUR	-0.44	0.0003	0.0000	-0.03	0.7313	-0.45	0.0003	0.0000	-0.04	0.7026
MCD	0.56	0.0015	0.0004	0.76	0.0000	0.57	0.0014	0.0003	0.78	0.0000
WHD	-0.80	0.0000	0.0000	-0.40	0.0040	-0.80	0.0000	0.0000	-0.37	0.0100
					4-Year-Ahead					
AE	-0.62	0.0000	0.0000	-0.20	0.0371	-0.61	0.0000	0.0000	-0.15	0.1083
EME	-0.34	0.0009	0.0001	-0.01	0.8863	-0.25	0.0179	0.0068	0.08	0.4302
LIC	-0.45	0.0000	0.0000	-0.31	0.0014	-0.44	0.0000	0.0000	-0.30	0.0019
AFR	-0.60	0.0000	0.0000	-0.50	0.0000	-0.60	0.0000	0.0000	-0.53	0.0000
APD	-0.28	0.0451	0.0144	0.09	0.4957	-0.22	0.1154	0.0528	0.22	0.0950
EUR	-0.54	0.0000	0.0000	-0.16	0.1245	-0.47	0.0001	0.0000	-0.10	0.3656
MCD	0.23	0.1960	0.1487	0.40	0.0218	0.36	0.0394	0.0188	0.63	0.0003
WHD	-0.79	0.0000	0.0000	-0.42	0.0021	-0.79	0.0000	0.0000	-0.39	0.0067
					3-Year-Ahead					
AE	-0.68	0.0000	0.0000	-0.28	0.0019	-0.62	0.0000	0.0000	-0.23	0.0120
EME	-0.37	0.0001	0.0000	-0.06	0.5157	-0.29	0.0027	0.0006	0.00	0.9709
LIC	-0.63	0.0000	0.0000	-0.50	0.0000	-0.59	0.0000	0.0000	-0.43	0.0000
AFR	-0.79	0.0000	0.0000	-0.72	0.0000	-0.77	0.0000	0.0000	-0.66	0.0000
APD	-0.32	0.0131	0.0030	0.11	0.3611	-0.19	0.1465	0.0822	0.21	0.0838
EUR	-0.56	0.0000	0.0000	-0.23	0.0294	-0.48	0.0000	0.0000	-0.21	0.0484
MCD	0.05	0.7759	0.7476	0.26	0.1217	0.10	0.5334	0.4746	0.33	0.0456
WHD	-0.82	0.0000	0.0000	-0.48	0.0002	-0.81	0.0000	0.0000	-0.44	0.0010

Source: Author's calculations using *WEO*.

Note that, at the five percent level, the statistically significant optimistic bias observed for the full sample holds, with few exceptions—MCD, using both vintages, and APD, for the fall vintage in three- and four-year-ahead forecasts—regardless of the development level and area departments.

C. Serial Correlation and General Efficiency

Informational efficiency requires not only unbiasedness (zero mean), but also absence of serial correlation (no correlation between past and current errors), and unpredictability of forecast errors (no other variable can be used to predict forecast errors).

Information about the *k*th-order serial correlation (ρ) in medium-term forecast errors, defined as the correlation between $e^{ij}_{t,t+k}$ and $e^{ij}_{t-k,t}$, is reported in Tables A2.1 and A2.2. Both the descriptive statistics and the regression-based statistics indicate that serial correlation of forecast errors may be restricted to about 20 percent of countries, when serial correlation is

tested independently from unbiasedness. However, the more restrictive hypothesis H_0: $\mu = \rho = 0$ is easier to reject. Thus, when unbiasedness and absence of serial correlation are jointly tested, around 40 percent of countries show evidence of serial correlation (Table A2.2). As in the case of unbiasedness, when possible small-sample distortions are taken into account using a bootstrap procedure, the associated *p*-values suggest that serial correlation may be much less frequent than indicated by standard *t*-tests.

More generally, if information about any variable x_t is available at the time of the forecast, x_t must not explain forecast errors. We follow Timmermann (2006) and test this hypothesis by estimating the following regression:

$$e^{ij}_{t,t+k} = \mu + \beta x_t + \varepsilon_t ,$$

where x_t is a variable that is observable at year *t*; efficiency is rejected if the estimated value of β is statistically different from zero. This method allows us to assess how well the IMF forecasts incorporate economic interdependencies among countries; the results are reported in Table 14 in the main text.

D. The Five-year Rule for the Closing of Output Gaps

Timmermann (2006) suggested that an informal rule used by IMF economists—which assumes that actual GDP converges to potential GDP within a period of five years—may be a source of optimistic biases.

For instance, consider a forecast made at time *t* for GDP at time *t*+5. Assume that, at time *t*, the forecaster both knows that GDP is below its trend (potential output) and correctly predicts the level of the GDP trend for time *t*+5. If the forecaster imposes the rule that the output gap is closed at *t*+5 or earlier, then the forecasted level of actual GDP at *t*+5 is forced to be very close to potential GDP and the forecasted output gap at *t*+5 is zero.

If, instead, actual GDP at *t*+5 is, say, five percent below the (correctly estimated) GDP trend—because the actual convergence of GDP, from below, towards its trend takes longer than five years—then there will be a systematic overprediction of GDP growth at all horizons *t*+*k*, for $0 \le k \le 5$. Moreover, the overprediction in the forecast error at *t*+5 of the GDP level, as a proportion of the forecasted value, should be also close to five percent.

In that case, for given paths of both potential and actual GDP between *t* and *t*+5, a forecast of zero for the output gap is associated with the largest possible (negative) forecast error for both the GDP level (i.e., five percent) and growth rate. If IMF economists violate the rule and predict a larger output gap, of say one percent below trend, then the forecast error in the GDP level as a percentage of the forecast would be less than the five percent (in absolute terms) obtained under the informal rule. The error on the growth rate would also be smaller in absolute terms.

If these situations are pervasive, when economists try to respect the informal rule smaller *predicted* output gaps in absolute terms (less negative) for year *t+k* should be associated with larger absolute forecast errors (more negative) for that year. The correlation between errors and the forecasted output gap should therefore be negative.[9] Results presented in Table 15 in the main text, suggest that these situations may indeed be occurring.

[9] The tendency for a negative correlation between forecast errors and forecasted output gap also occurs if convergence to the trend from above takes longer than five years. However, in that case, forecast errors would be positive (pessimistic forecasts)—which is not consistent with the evidence. To satisfy the informal rule of zero output gap at *t*+5 and still be optimistic, the forecaster would have to adjust the estimated level of potential GDP *upwards* (rather than actual GDP downwards). In that case, the correlation switches to positive or null. This may partially explain why the negative correlations in Table 15 in the main text, while statistically significant, are not too high.

ANNEX 3. STATISTICAL ANALYSIS OF BIASES IN FORECASTS IN THE CONTEXT OF IMF PROGRAMS

The comparative analysis of biases in the context of IMF programs relative to no-program situations is based on pooled least squares (PLS) estimations of variants of equation (A2.3), discussed in Annex 2, where dummy variables are used to split the sample into different categories of interest. Results are displayed in Tables A3.1-A3.4 and their main features summarized in Figure A3.1.[1]

A. Program versus No Program (At the Time the Forecast is Made)

The results reported in Table A3.1 and shown in Boxes (B) and (C) of Figure A3.1 follow from replacing the constant term in equation (A2.3) by two dummy variables. The first dummy variable (D_t^{NoProg}) refers to the "no-program case" and takes the value of one when an observation about a forecast made at year *t* refers to a country that is not in an IMF program at that time, and zero otherwise. The second dummy variable (D_t^{Prog}) refers to the "program case" and equals one when $D_t^{NoProg} = 0$ (that is, when the country is in an IMF program at year *t*).[2] The estimated equation becomes:

$$e_{t,t+k}^{ij}(h) = \mu^{NoProg} D_t^{NoProg} + \mu^{Prog} D_t^{Prog} + \varepsilon_t(h). \qquad \text{(A3.1)}$$

Table A3.1. IMF Programs (I): Average bias in GDP growth forecasts in PLS regressions

	Spring, 2 year window					Fall, 2 year window				
		p-value		"Big" Recessions			*p*-value		"Big" Recessions	
Sub sample	Bias	Pooled LS	FE	Bias	*p*-value	Bias	Pooled LS	FE	Bias	*p*-value
5-Year-Ahead										
No program at time *t*	-0.33	0.0000	0.0000	0.00	0.9722	-0.36	0.0000	0.0000	-0.02	0.7560
Program at time *t*	-0.34	0.0048	0.0009	-0.08	0.4524	-0.27	0.0206	0.0087	-0.09	0.4291
Statistically different?	No			No		No			No	
4-Year-Ahead										
No program at time *t*	-0.46	0.0000	0.0000	-0.16	0.0214	-0.37	0.0000	0.0000	-0.06	0.4068
Program at time *t*	-0.39	0.0009	0.0001	-0.17	0.1312	-0.45	0.0002	0.0000	-0.24	0.0417
Statistically different?	No			No		No			Yes	
3-Year-Ahead										
No program at time *t*	-0.50	0.0000	0.0000	-0.23	0.0004	-0.41	0.0000	0.0000	-0.13	0.0392
Program at time *t*	-0.62	0.0000	0.0000	-0.42	0.0002	-0.66	0.0000	0.0000	-0.45	0.0001
Statistically different?	No			No		Yes			Yes	

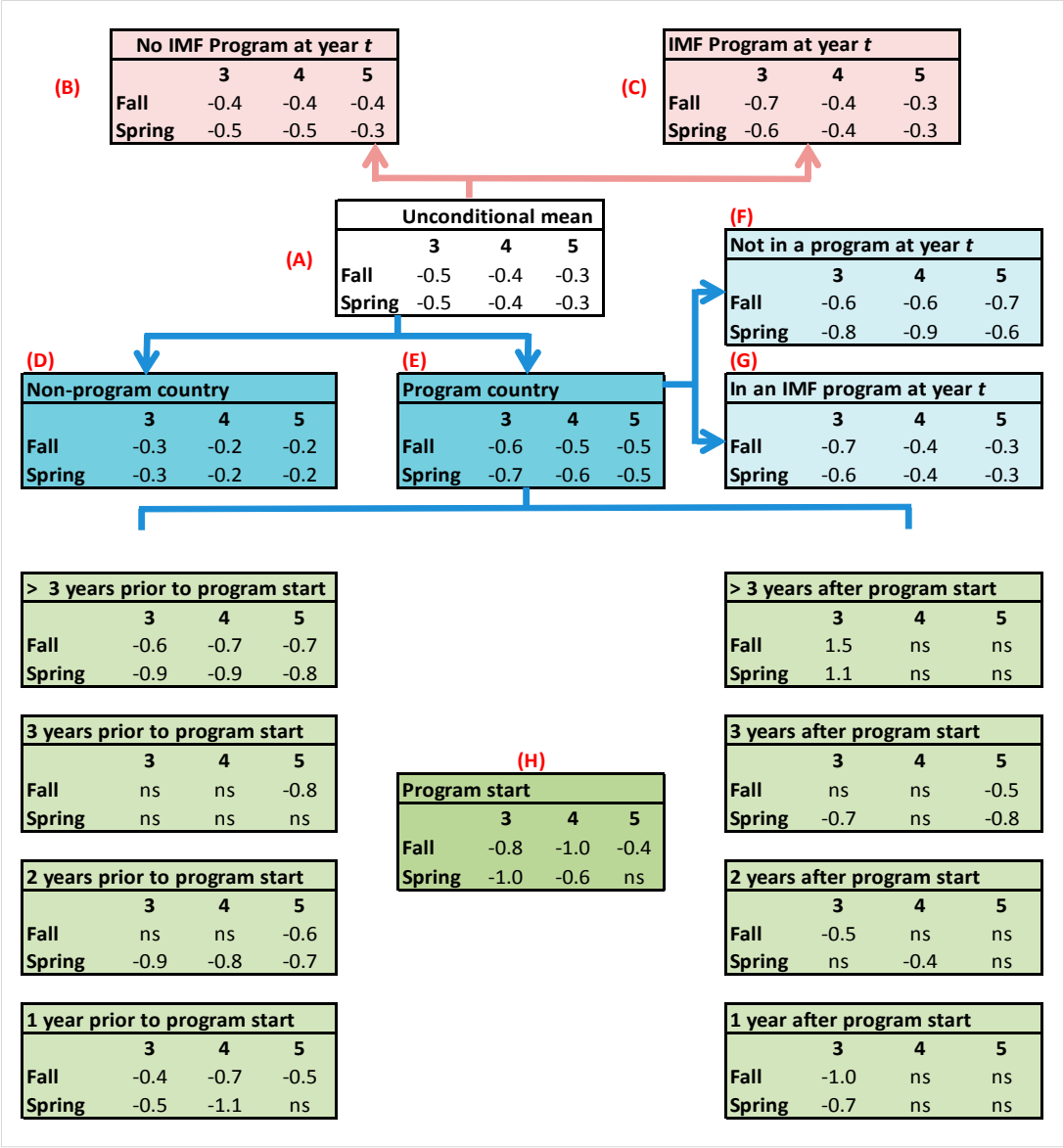
Source: Author's calculations using *WEO*.

[1] Figure A3.1highlights the statistically significant results, and refers to biases in the baseline estimations without controlling for "big recessions."

[2] This exercise is equivalent to estimating equation (A2.3) in two subsamples consistent with the two dummy variables.

Note that because $D_t^{NoProg} + D_t^{Prog} = 1$, and all observations used to estimate equation (A2.3) are being considered in one case or the other, no information is lost and the weighted average of coefficients μ^{NoProg} and μ^{Prog} is equal to the estimated value of μ in (A2.3).[3] These coefficients will provide estimates of the bias in each case considered.

Figure A3.1. Bias in medium-term forecasts in the context of programs (made at year *t* for year *t+k*)
(in percentage points)



Note: "ns" denotes not statistically significant estimates. The results shown refer to biases in the baseline estimations for all vintages and horizons, i.e., without controlling for "big recessions."

Source: Author's calculations using *WEO*.

[3] The weights are the ratios of observations in one particular case (no-program or program) to the total number of observations used to estimate (A2.3).

As discussed in the main text, the mean forecast errors are found to be always negative and statistically significant regardless of whether forecasts are made during IMF programs or not. Therefore, *optimistic biases do exist whether countries are in program status or not*. However, there is little evidence of statistically significant differences between program and no-program cases, except in three-year-ahead forecasts. The evidence in favor of any additional program-related optimistic bias is weak.

The results in columns 5-6 and 10-11 of Table A3.1 provide somewhat stronger evidence of additional program-related optimistic bias. The sizeable reduction in optimistic biases when controlling for big recessions is also observed when the sample is split between program- and no-program observations. However, for three- and four-year-ahead forecasts, the dummy variables for big recessions do not completely wipe out the differences between program and no-program cases. On the contrary, the reduction in bias is much more pronounced among no-program observations—which makes the larger optimistic bias in programs relative to normal times more noticeable, especially in three- and four-year-ahead forecasts.

B. Program Countries *vs.* Non-Program Countries

Perhaps the forecast errors associated with countries with a higher likelihood of seeking IMF programs may be drawn from a different probability distribution than the forecast errors for countries with no history of programs. To investigate this possibility, observations are sorted according to countries' history of IMF program participation. Results are shown in Table A3.2 and Figure A3.1 (Boxes D and E). The estimated PLS equation becomes:

$$e^{\,ij}_{t,t+k}(h) = \mu^{NoPC} D_t^{NoPC} + \mu^{PC} D_t^{PC} + \varepsilon_t(h). \qquad \text{(A3.2)}$$

Again, the constant term in (A2.3), from Annex 2, is replaced by two dummy variables. The first dummy variable (D_t^{NoPC}) takes the value of one when a particular observation refers to a country that has never had an IMF program during the sample period (and zero, otherwise). We refer to this as the "non-program country" case. The second dummy (D_t^{PC}) equals one when the observation at year *t*, the year the forecast is made, comes from a country that has had an IMF program at some point during the sample period (although not necessarily at year *t*). We refer to this case as the "program country" case.

Especially because the difference in the estimated bias between program and no-program observations is substantially larger in equation (A3.2) than in equation (A3.1), the comparison of results from these two estimated equations—reported in Tables A3.1 and A3.2, respectively—indicates that some of the observations counted as "no-program" in the estimation of (A3.1)—more specifically, those coming from program countries that are not in an IMF program at the time of the forecast—are associated with larger biases. Part of the additional optimistic bias in program countries, therefore, may be driven by factors that are specific to countries that seek IMF programs, and not by the program status itself.

Table A3.2. IMF programs (ii): Average bias in GDP growth forecasts in PLS regressions

	Spring, 2-year window					Fall, 2-year window				
		p-value		"Big" Recessions			*p*-value		"Big" Recessions	
Sub sample	Bias	Pooled LS	FE	Bias	*p*-value	Bias	Pooled LS	FE	Bias	*p*-value
5-Year-Ahead										
"Non-program country"	-0.15	0.0917	0.0399	0.21	0.0092	-0.16	0.0917	0.0420	0.23	0.0083
"Program country"	-0.48	0.0000	0.0000	-0.22	0.0085	-0.47	0.0000	0.0000	-0.21	0.0102
4-Year-Ahead										
"Non-program country"	-0.18	0.0469	0.0187	0.15	0.0562	-0.20	0.0264	0.0090	0.13	0.1032
"Program country"	-0.65	0.0000	0.0000	-0.42	0.0000	-0.53	0.0000	0.0000	-0.30	0.0003
3-Year-Ahead										
"Non-program country"	-0.28	0.0007	0.0001	0.01	0.8822	-0.30	0.0006	0.0001	0.01	0.9269
"Program country"	-0.73	0.0000	0.0000	-0.51	0.0000	-0.60	0.0000	0.0000	-0.38	0.0000

Source: Author's calculations using *WEO*.

To further investigate this matter, the observations associated with program countries in (A3.2) are split into two sub-categories. This is reflected in Figure A3.1 by the arrows departing from Box (E) to Boxes (F) and (G) and is also reported in Table A3.3. At year *t*, when forecasts are made, program countries may or may not be in an IMF program and this information is captured by dummy variables $D_t^{PC(t)}$ and $D_t^{PC(not\ t)}$, respectively. The estimated equation becomes:

$$e^{\ ij}_{t,t+k}(h) = \mu^{NoPC} D_t^{NoPC} + \mu^{PC(not\ t)} D_t^{PC(not\ t)} + \mu^{PC(t)} D_t^{PC(t)} + \varepsilon_t(h). \qquad \text{(A3.3)}$$

Again, equation (A3.3) is estimated with no loss of information relative to (A2.3). Because $D_t^{NoPC} + D_t^{PC(not\ t)} + D_t^{PC(t)} = 1$ and $D_t^{PC(not\ t)} + D_t^{PC(t)} = D_t^{PC}$, the weighted average of coefficients μ^{NoPC}, $\mu^{PC(not\ t)}$, and $\mu^{PC(t)}$ equals the estimated value of μ in (A2.3), and the weighted average of coefficients $\mu^{PC(not\ t)}$ and $\mu^{PC(t)}$ is equal to the estimated value of μ^{PC} in (A3.2). If no dummy variables are used (for big recessions), the estimated values of μ^{NoPC} obtained in the estimations of (A3.2) and (A3.3) are identical.

The results in Table A3.3 provide a nuanced characterization of program-related optimistic biases: (i) optimistic biases in program countries, whether in an IMF program at year *t* or not, are always larger than those in non-program countries; (ii) forecasts made in program countries during programs are always more optimistically biased than forecasts in non-program countries, but they are *less* optimistic than in program countries that are *not* in a program at the time of the forecast.

Table A3.3. IMF programs (III): Average bias in GDP growth forecasts in PLS regressions

	Spring, 2 year window					Fall, 2 year window				
		p-value		"Big" Recessions			*p*-value		"Big" Recessions	
Sub sample	Bias	Pooled LS	FE	Bias	p-value	Bias	Pooled LS	FE	Bias	p-value
5-Year-Ahead										
"Non-program country"	-0.15	0.0917	0.0399	0.21	0.0092	-0.16	0.0917	0.0420	0.23	0.0083
"Program country"	-0.48	0.0000	0.0000	-0.22	0.0085	-0.47	0.0000	0.0000	-0.21	0.0102
Not at time *t*	-0.61	0.0000	0.0000	-0.33	0.0057	-0.65	0.0000	0.0000	-0.38	0.0015
at time *t*	-0.34	0.0048	0.0009	-0.08	0.4524	-0.27	0.0207	0.0087	-0.09	0.4291
4-Year-Ahead										
"Non-program country"	-0.18	0.0469	0.0187	0.15	0.0562	-0.20	0.0264	0.0090	0.13	0.1032
"Program country"	-0.65	0.0000	0.0000	-0.42	0.0000	-0.53	0.0000	0.0000	-0.30	0.0003
Not at time *t*	-0.88	0.0000	0.0000	-0.62	0.0000	-0.61	0.0000	0.0000	-0.32	0.0052
at time *t*	-0.39	0.0009	0.0001	-0.17	0.1312	-0.45	0.0002	0.0000	-0.24	0.0417
3-Year-Ahead										
"Non-program country"	-0.28	0.0007	0.0001	0.01	0.8822	-0.30	0.0006	0.0001	0.01	0.9269
"Program country"	-0.73	0.0000	0.0000	-0.51	0.0000	-0.60	0.0000	0.0000	-0.38	0.0000
Not at time *t*	-0.82	0.0000	0.0000	-0.58	0.0000	-0.56	0.0000	0.0000	-0.32	0.0029
at time *t*	-0.62	0.0000	0.0000	-0.42	0.0002	-0.66	0.0000	0.0000	-0.45	0.0001

Source: Author's calculations using *WEO*.

Timing of optimistic biases during programs

The analysis based on equation (A3.3) above treats equally all observations related to program years. It is possible, however, that within program countries during a program, the optimistic bias, if any, changes in different stages of the program or even pre-dates the program and is carried over to program years.

To analyze the timing of optimistic biases in program countries around program dates, the observations from the subsample of program countries are split according to the year that forecasts are made relative to when programs start. That is, instead of classifying observations specific to program countries according to participation in a program or not, they are divided into nine categories, each representing one year, starting four years or more before, and ending four years or more after, programs start, at year T_0.

Let T_0, T_0-z, and T_0+z represent observations from the year the program starts, z years prior, and z years after program start, respectively. The dummy variable D_t^{PC}, for program countries (regardless whether engaged in programs at time t or not) can be split into eight "timing" dummy variables: six for T_0-z and T_0+z, with z = 1, 2, 3, and two others to capture observations from periods four or more years before ($< T_0$-3) or after ($> T_0$+3) program start. For $0 \leq z \leq 3$, all observations belong to program years. Some (very few) observations associated with the dummy variable for $t > T_0$+3 also belong to program years. As in other cases discussed above, there is no lost information relative to equations A2.3 and A3.1−3.

The results in Table A3.4 follow from the PLS estimation of a version of equation (A2.3) in which these timing-indicative dummy variables are used. In Figure A3.1, biases in forecasts made in the years preceding and following programs are represented in the green boxes to the

left and to the right of Box (H), respectively. Box (H) represents biases in forecasts made at the year programs start.

Table A3.4. IMF programs (IV): Average bias in GDP growth forecasts in PLS regressions

	Spring, 2 year window					Fall, 2 year window				
		p-value		"Big" Recessions			*p*-value		"Big" Recessions	
Sub sample	Bias	Pooled LS	FE	Bias	p-value	Bias	Pooled LS	FE	Bias	p-value
5-Year-Ahead										
"Non-program country"	-0.15	0.0917	0.0399	0.21	0.0092	-0.16	0.0917	0.0420	0.23	0.0083
"Program country"	-0.48	0.0000	0.0000	-0.22	0.0085	-0.47	0.0000	0.0000	-0.21	0.0102
< T_0 - 3	-0.82	0.0000	0.0000	-0.48	0.0021	-0.72	0.0000	0.0000	-0.35	0.0289
T_0 - 3	-0.32	0.4045	0.3197	-0.14	0.7201	-0.75	0.0592	0.0420	-0.67	0.0979
T_0 - 2	-0.70	0.0645	0.0501	-0.54	0.1412	-0.58	0.0971	0.0930	-0.31	0.3500
T_0 - 1	-0.10	0.6946	0.7077	0.22	0.3893	-0.47	0.0795	0.0870	-0.27	0.3063
Program start (T_0)	-0.23	0.3021	0.2597	0.10	0.6378	-0.40	0.0641	0.0632	-0.08	0.7074
T_0 + 1	-0.26	0.2125	0.2218	0.08	0.6969	-0.06	0.7706	0.7671	0.22	0.2958
T_0 + 2	-0.36	0.1330	0.1172	-0.12	0.5941	-0.29	0.1977	0.2027	0.04	0.8559
T_0 + 3	-0.81	0.0169	0.0124	-0.46	0.1716	-0.53	0.0893	0.1161	-0.21	0.5536
> T_0 + 3	-0.26	0.6873	0.6620	-0.20	0.7533	0.08	0.9269	0.9332	0.26	0.7578
4-Year-Ahead										
"Non-program country"	-0.18	0.0469	0.0187	0.15	0.0562	-0.20	0.0264	0.0090	0.13	0.1032
"Program country"	-0.65	0.0000	0.0000	-0.42	0.0000	-0.53	0.0000	0.0000	-0.30	0.0003
< T_0 - 3	-0.89	0.0000	0.0000	-0.59	0.0001	-0.65	0.0000	0.0000	-0.30	0.0506
T_0 - 3	-0.32	0.4554	0.6093	-0.23	0.5877	-0.30	0.4926	0.6009	-0.13	0.7677
T_0 - 2	-0.80	0.0140	0.0118	-0.54	0.0848	-0.48	0.1535	0.1986	-0.21	0.5293
T_0 - 1	-1.08	0.0002	0.0007	-0.89	0.0017	-0.65	0.0103	0.0120	-0.45	0.0694
Program start (T_0)	-0.64	0.0029	0.0039	-0.38	0.0796	-0.95	0.0000	0.0000	-0.75	0.0004
T_0 + 1	-0.27	0.2140	0.1997	0.02	0.9252	-0.21	0.3598	0.3500	0.08	0.7046
T_0 + 2	-0.40	0.0655	0.0645	-0.04	0.8696	-0.27	0.2145	0.1781	0.01	0.9709
T_0 + 3	-0.20	0.5562	0.4986	0.05	0.8821	-0.12	0.7314	0.7378	0.11	0.7563
> T_0 + 3	0.23	0.7009	0.7086	0.29	0.6387	0.79	0.2857	0.4004	0.79	0.2859
3-Year-Ahead										
"Non-program country"	-0.28	0.0007	0.0001	0.01	0.8822	-0.30	0.0006	0.0001	0.01	0.9269
"Program country"	-0.73	0.0000	0.0000	-0.51	0.0000	-0.60	0.0000	0.0000	-0.38	0.0000
< T_0 - 3	-0.92	0.0000	0.0000	-0.61	0.0000	-0.65	0.0000	0.0000	-0.30	0.0301
T_0 - 3	-0.58	0.1209	0.2046	-0.52	0.1584	-0.60	0.2160	0.3168	-0.47	0.3222
T_0 - 2	-0.93	0.0079	0.0129	-0.77	0.0248	-0.26	0.4041	0.3912	-0.05	0.8576
T_0 - 1	-0.51	0.0400	0.0741	-0.30	0.2155	-0.40	0.0943	0.1086	-0.24	0.2962
Program start (T_0)	-0.99	0.0000	0.0000	-0.77	0.0003	-0.77	0.0001	0.0001	-0.58	0.0017
T_0 + 1	-0.69	0.0010	0.0012	-0.45	0.0299	-0.95	0.0000	0.0000	-0.73	0.0006
T_0 + 2	-0.32	0.1808	0.1587	-0.02	0.9411	-0.52	0.0383	0.0417	-0.20	0.4205
T_0 + 3	-0.65	0.0412	0.0393	-0.38	0.2168	-0.21	0.5044	0.5192	0.07	0.8180
> T_0 + 3	1.10	0.0844	0.1466	1.15	0.0710	1.47	0.0892	0.0371	1.47	0.0893

Source: Author's calculations using *WEO*.

We find that additional optimistic biases relative to non-program situations are mostly found in the year that program start. Statistically significant optimistic biases, with one exception (five-year-ahead forecasts, Fall vintage), disappear at T_0-3 but are often present at two years and one year before program start, although they are smaller (in absolute size) than the biases in forecasts made before T_0-3. In one case (four-year-ahead forecasts, Spring vintage), the bias remains high—above one percentage point—up to just one year before program start.

At program start, the average biases in four-year-ahead (Fall vintage) and three-year-ahead forecasts (both vintages) increase again (Figure A3.1, Box H).[4] Moreover, the optimistic biases in forecasts made at T_0 are always larger than the average bias for non-program countries (Figure A3.1, Box D). Relative to program countries that are not in a program at time *t* (Box F), the optimism in forecasts made at T_0 is statistically larger in three-year-ahead forecasts, but the evidence is mixed for four-year-ahead forecasts (only true for the Fall vintage) and nonexistent in five-year-ahead forecasts.

The larger optimistic bias in forecasts that are made the year that programs start, relative to non-program cases, usually disappears one year after (see Figure A3.1, boxes to the right of H), with the exception of three-year-ahead forecasts. In most cases, the optimistic bias fades out in the subsequent years after the program starts, though it sometimes reappears. Bias in four- and five-year-ahead forecasts recurs in T_0+2 (Spring vintage only) and T_0+3, respectively. In the case of three-year-ahead forecasts, the bias eventually reverts to a *pessimistic* bias, mainly reflecting forecasts made in 2008, after the start of the recent financial crisis, which underestimated the 2010 recovery.

Type of program

It is also possible that the type of program matters for biases. To investigate this possibility, we estimate the following equation:

$$e^{\,ij}_{t,t+k}(h) = \mu^{NoPC} D_t^{NoPC} + \mu^{PC(not\ t)} D_t^{PC(not\ t)} + \mu^{SBA} D_t^{SBA} + \mu^{PRGF} D_t^{PRGF} + \varepsilon_t(h), \tag{A3.4}$$

where the dummy variable $D_t^{PC(t)}$ in (A3.3) is further split between D_t^{SBA} and D_t^{PRGF}, which represent observations from program countries that are in SBA-type and PRGF-type programs, respectively. The former are lending facilities funded by the IMF General Resources Account (GRA) and include Stand-by (SBA), Precautionary Credit Line (PCL), Extended Fund Facility (EFF), and first credit tranche (FCTA) arrangements. The latter are concessional lending facilities created under the Poverty Reduction and Growth Trust (PRGT), which includes programs under the Exogenous Shocks Facility (ESF), Structural Adjustment Facility (SAF), Enhanced Structural Adjustment Facility (ESAF), Extended Credit Facility (ECF), and Standby Credit Facility (SCF).

The results, displayed in Table A3.5, indicate that SBA-type programs show larger optimistic biases relative to PRGF-type programs in the four- and, especially, the three-year-ahead forecasts. However, the differences are not statistically significant.

[4] This occurs regardless of the introduction of dummy variables for big recessions.

Table A3.5. IMF programs (V): Average bias in GDP growth forecasts in PLS regressions

	Spring, 2 year window					Fall, 2 year window				
		p-value		"Big" Recessions			*p*-value		"Big" Recessions	
Sub sample	bias	Pooled LS	FE	bias	p-value	bias	Pooled LS	FE	bias	p-value
5-Year-Ahead										
SBA	-0.14	0.4919	0.3964	0.19	0.3265	-0.35	0.0822	0.0336	-0.01	0.9614
PRGF	-0.43	0.0028	0.0005	-0.21	0.1232	-0.23	0.1155	0.0777	-0.04	0.7899
4-Year-Ahead										
SBA	-0.51	0.0138	0.0017	-0.21	0.2813	-0.54	0.0092	0.0021	-0.27	0.1940
PRGF	-0.33	0.0203	0.0084	-0.14	0.3043	-0.39	0.0068	0.0017	-0.21	0.1287
3-Year-Ahead										
SBA	-0.70	0.0010	0.0001	-0.41	0.0402	-0.80	0.0001	0.0000	-0.50	0.0106
PRGF	-0.59	0.0000	0.0000	-0.42	0.0026	-0.58	0.0000	0.0000	-0.42	0.0021

Source: Author's calculations using *WEO*.

ANNEX 4. STATISTICAL ANALYSIS OF RESPONSES TO THE SURVEY QUESTION ABOUT IMF MEDIUM-TERM FORECASTS

This annex provides a more detailed account of the responses to the evaluation survey question on medium-term forecasts. It reports on statistical comparisons of responses from country authorities, private sector representatives, and staff, as well as on differences across groups of countries.

Survey respondents were asked to express their general agreement or disagreement with the following series of statements, or prompts, covering different aspects of medium-term forecasting activities at the Fund:

(a) "The IMF medium-term forecasts for my country are based on reasonable conditioning assumptions about world economic conditions;"

(b) "For policy discussions, medium-term forecasts are more valuable than one or two-year forecasts;"

(c) "In its analysis, the IMF should place less emphasis on one and two-year forecasts and more emphasis on medium-term forecasts;"

(d) "Medium-term forecasts entail too much uncertainty to be valuable for policy discussions;"

(e) "When making medium-term forecasts, it is reasonable to assume that my country's output gap is closed by the end of five years;"

(f) "When making medium-term forecasts, it is reasonable to assume that my country's real exchange rate remains fixed for five years."

The following additional prompt was included in the survey sent to country authorities and the private sector:

(g) "We largely ignore the medium-term forecasts."

For each prompt, respondents were given six possible alternative answers to choose from. The quantitative analysis uses indices, or response scores, based on averaging the responses according to the following mapping:

Response scores

Answer	Value
"Not applicable / I do not know"	0
"Strongly disagree"	1
"Disagree"	2
"Neither agree nor disagree"	3
"Agree"	4
"Strongly agree"	5

Table A4.1 summarizes the survey responses, based on the aggregated scores, and formally compares the desk economists' and authorities' responses to prompts (a)–(f).[1] Tables A4.2–4 and A4.5(a)–(b) present scores from the survey results disaggregated between program and non-program countries (Table A4.2) and among countries at different levels of economic development, or income (Table A4.3) and covered by different IMF area departments (Tables A4.4 and A4.5(a)–(b)). All these tables show tests of equality of means: *t*-tests to compare responses from two subsamples and/or *F*-tests to assess the equality of average responses across three or more subsamples.

Table A4.1. Responses by IMF desk economists and country authorities: Summary statistics

	IMF Desk Economists			Authorities			
Question	Mean	Std	Skewness	Mean	Std	Skewness	Equal means? (1)
(a)	3.97	0.80	-1.08	3.87	0.52	-1.18	0.1101
(b)	2.77	0.94	0.18	3.06	0.94	0.26	0.0047
(c)	2.31	0.84	0.89	2.70	0.85	0.50	0.0000
(d)	3.06	1.07	0.17	3.05	0.95	0.14	0.4862
(e)	3.07	1.07	-0.38	3.16	0.91	-0.32	0.2325
(f)	2.85	1.11	-0.02	2.81	1.02	0.19	0.3704

(1) *p*-value associated with the null hypothesis of equal means in a one-sided *t*-test, assuming unequal variances. Results do not qualitatively change when assuming equal variances.

Source: Author's calculations using IEO Forecast evaluation survey.

Table A4.2. Survey results: Program vs. non-program countries

Question	Full sample			Program			Non-Program			Reject
	Mean	Std	Skewness	Mean	Std	Skewness	Mean	Std	Skewness	equal means? (1)
Country authorities										
(a)	3.87	0.52	-1.18	3.98	0.61	-1.32	3.83	0.49	-1.23	0.0885
(b)	3.06	0.94	0.26	3.24	1.03	0.26	2.99	0.90	0.21	0.0738
(c)	2.70	0.85	0.50	2.93	0.93	0.49	2.62	0.81	0.43	0.0263
(d)	3.05	0.95	0.14	3.39	1.02	-0.02	2.94	0.91	0.12	0.0059
(e)	3.16	0.91	-0.32	3.32	1.04	-0.66	3.11	0.86	-0.20	0.1215
(f)	2.81	1.02	0.19	2.74	1.00	0.24	2.83	1.03	0.16	0.3178
(g)	2.33	0.85	0.54	2.11	1.01	0.73	2.40	0.78	0.61	0.0398
IMF desk economists										
(a)	3.97	0.80	-1.08	4.07	0.74	-2.00	3.94	0.82	-2.48	0.2098
(b)	2.77	0.94	0.18	2.80	0.92	0.26	2.76	0.95	-0.35	0.4198
(c)	2.31	0.84	0.89	2.27	0.83	0.00	2.32	0.85	-0.16	0.3742
(d)	3.06	1.07	0.17	3.07	1.01	-0.44	3.05	1.09	-0.29	0.4732
(e)	3.07	1.07	-0.38	3.20	1.13	-0.89	3.03	1.06	-0.90	0.2357
(f)	2.85	1.11	-0.02	2.87	1.20	-0.16	2.84	1.09	-0.30	0.4629

(1) *p*-value associated with the null hypothesis of equal means in one-sided *t*-test, assuming unequal variances.

Source: Author's calculations using IEO Forecast evaluation survey.

[1] In comparisons between the authorities' and staff's responses, respondents who answered "not applicable/do not know" are not considered.

Table A4.3. Survey results: Test of equality of means by level of development

Question	Mean				Reject equal means? (1)			
	Full sample	Advanced	Emerging	Low-Income	AE *vs.* EME	AE *vs.* LIC	EME *vs.* LIC	*F*-test
Country Authorities								
(a)	3.87	3.87	3.83	3.92	0.3459	0.2923	0.1761	0.6484
(b)	3.06	2.87	3.09	3.17	0.0883	0.0546	0.3333	0.2243
(c)	2.70	2.53	2.73	2.81	0.0000	0.0380	0.3204	0.1699
(d)	3.05	2.81	3.11	3.19	0.0382	0.0193	0.3167	0.0786
(e)	3.16	3.34	2.99	3.23	0.0190	0.2896	0.0750	0.0898
(f)	2.81	2.77	2.91	2.71	0.2239	0.3815	0.1429	0.5418
(g)	2.33	2.67	2.20	2.21	0.0038	0.0059	0.4790	0.0132
IMF Desk Economists								
(a)	3.97	3.90	4.11	3.79	0.1778	0.3197	0.0218	0.1229
(b)	2.77	2.70	2.77	2.81	0.3925	0.3335	0.4081	0.9106
(c)	2.31	2.40	2.31	2.26	0.3627	0.2938	0.3757	0.8550
(d)	3.06	3.35	2.83	3.26	0.0394	0.3841	0.0173	0.0583
(e)	3.07	2.65	3.17	3.12	0.0445	0.0673	0.3981	0.2191
(f)	2.85	2.55	2.80	3.07	0.2067	0.0554	0.1071	0.2343

(1) *p*-values associated with the null hypothesis of equal means in one-sided *t*-tests (pairwise tests) and a *F*-test (joint test), assuming unequal variances.

Source: Author's calculations using IEO Forecast evaluation survey.

Table A4.4. Survey results: Test of equality of means by area departments—F-test

Question	Mean						*F*-test (1)
	Full sample	AFR	APD	EUR	MCD	WHD	
Country Authorities							
(a)	3.83	3.93	3.93	3.88	3.82	3.78	0.7909
(b)	2.99	3.19	3.07	2.89	3.23	3.08	0.5494
(c)	2.62	2.68	2.62	2.63	2.71	2.89	0.7224
(d)	2.94	3.55	3.03	3.00	2.90	2.81	0.0337
(e)	3.11	3.31	3.07	3.22	3.22	3.00	0.7118
(f)	2.83	2.72	2.48	2.85	3.35	2.77	0.0815
(g)	2.40	2.23	2.60	2.47	2.05	2.14	0.0499
IMF Desk Economists							
(a)	3.97	3.81	3.88	4.00	3.77	4.29	0.0258
(b)	2.77	2.67	2.82	2.78	2.41	3.11	0.1401
(c)	2.31	2.19	2.59	2.25	2.14	2.46	0.2994
(d)	3.06	3.19	2.88	3.31	3.18	2.64	0.1203
(e)	3.07	2.89	3.29	2.91	3.41	3.04	0.2819
(f)	2.85	3.15	2.65	2.84	2.95	2.61	0.4684

(1) *p*-value associated with the null hypothesis of equal means in a *F*-test, assuming unequal variances.

Source: Author's calculations using IEO Forecast evaluation survey.

Table A4.5(a). Survey results: Test of equality of means by area departments—pairwise t-tests [1]

Country Authorities									
(a)	*vs.* APD	*vs.* EUR	*vs.* MCD	*vs.* WHD	(b)	*vs.* APD	*vs.* EUR	*vs.* MCD	*vs.* WHD
AFR	0.5000	0.3267	0.2584	0.1470	AFR	0.3011	0.0731	0.4564	0.3148
APD		0.3077	0.2495	0.1321	APD		0.2015	0.3021	0.4711
EUR			0.3435	0.1979	EUR			0.1196	0.1581
MCD				0.4103	MCD				0.3147
(c)	*vs.* APD	*vs.* EUR	*vs.* MCD	*vs.* WHD	(d)	*vs.* APD	*vs.* EUR	*vs.* MCD	*vs.* WHD
AFR	0.3956	0.4005	0.4433	0.1675	AFR	0.0205	0.0053	0.0126	0.0012
APD		0.4764	0.3595	0.1133	APD		0.4373	0.3240	0.1700
EUR			0.3621	0.0883	EUR			0.3504	0.1619
MCD				0.2543	MCD				0.3570
(e)	*vs.* APD	*vs.* EUR	*vs.* MCD	*vs.* WHD	(f)	*vs.* APD	*vs.* EUR	*vs.* MCD	*vs.* WHD
AFR	0.1735	0.3491	0.3836	0.1186	AFR	0.1832	0.2972	0.0221	0.4279
APD		0.2213	0.2773	0.3855	APD		0.0556	0.0028	0.1291
EUR			0.4933	0.1401	EUR			0.0353	0.3622
MCD				0.2027	MCD				0.0268
(g)	*vs.* APD	*vs.* EUR	*vs.* MCD	*vs.* WHD					
AFR	0.0458	0.1016	0.2095	0.3108					
APD		0.2724	0.0109	0.0098					
EUR			0.0240	0.0206					
MCD				0.3169					

(1) *p*-value associated with the null hypothesis of equal means in one-sided *t*-tests, assuming unequal variances.

Source: Author's calculations using IEO Forecast evaluation survey.

Table A4.5(b). Survey results: Test of equality of means by area departments—pairwise t-tests [1]

IMF Desk Economist									
(a)	*vs.* APD	*vs.* EUR	*vs.* MCD	*vs.* WHD	(b)	*vs.* APD	*vs.* EUR	*vs.* MCD	*vs.* WHD
AFR	0.3742	0.2253	0.4396	0.0183	AFR	0.3061	0.3075	0.1470	0.0401
APD		0.2454	0.3051	0.0016	APD		0.4455	0.1013	0.1921
EUR			0.1811	0.0591	EUR			0.0658	0.0962
MCD				0.0135	MCD				0.0052
(c)	*vs.* APD	*vs.* EUR	*vs.* MCD	*vs.* WHD	(d)	*vs.* APD	*vs.* EUR	*vs.* MCD	*vs.* WHD
AFR	0.0580	0.3859	0.3922	0.1034	AFR	0.1641	0.3382	0.4953	0.0267
APD		0.1150	0.0371	0.3267	APD		0.0905	0.1599	0.2057
EUR			0.3003	0.1977	EUR			0.3295	0.0113
MCD				0.0638	MCD				0.0235
(e)	*vs.* APD	*vs.* EUR	*vs.* MCD	*vs.* WHD	(f)	*vs.* APD	*vs.* EUR	*vs.* MCD	*vs.* WHD
AFR	0.0932	0.4762	0.0347	0.3130	AFR	0.0798	0.1610	0.3879	0.0481
APD		0.1051	0.3497	0.2078	APD		0.2714	0.3274	0.4528
EUR			0.0412	0.3355	EUR			0.4339	0.2066
MCD				0.1051	MCD				0.3031

(1) *p*-value associated with the null hypothesis of equal means in one-sided *t*-tests, assuming unequal variances.

Source: Author's calculations using IEO Forecast evaluation survey.

Items 1 to 7 below analyze, in turn, the answers from country authorities' and desk economists to prompts (a)-(g) in the survey question about medium-term forecasts.

1. Desk economists, country authorities, and representatives from the private sector largely think that the medium-term forecasts for "their" countries are based on reasonable underlying assumptions about the global economy.

A large proportion, about 80 percent, of both desk economists and authorities either agree or strongly agree with prompt (a): "The IMF medium-term forecasts for my country are based on reasonable conditioning assumptions about world economic conditions." The distributions of responses to this prompt are strongly skewed to the left, suggesting overall agreement with the statement (Table A4.1). The average responses in both samples are not statistically different at the 10 percent significance level—see last column of Table A4.1.

The strong support for the underlying assumptions in medium-term IMF forecasts is robust to different partitions of the data from both staff and authorities, across program and non-program countries (Table A4.2), different levels of country development (Table A4.3), or across different area departments (Tables A4.4 and A4.5 (a)–(b)). With only three exceptions,[2] the tests of equality of means in Tables A4.1 and A4.2–A4.4 do not reject the hypothesis of equal means at standard significance levels, giving a clear indication of widespread support for the global assumptions used in medium-term forecasts.

The paucity of evidence of an organizational bias (whereby support from staff to prompt (a) would be stronger than support from the authorities) already mentioned for the full sample, is also seen when the data are partitioned by country development level and IMF area department. Although staff response scores to prompt (a) are higher (although not is a statistically significant sense) than those of the authorities (Table A4.1), the same does not hold for LICs (Table A4.3) nor for countries covered by the African (AFR), Asia and Pacific (APD), or Middle East and Central Asia (MCD) departments (Table A4.4).

2. While country authorities seem neutral on the matter, desk economists tend to value medium-term forecasts less than short-term forecasts for policy discussion purposes.

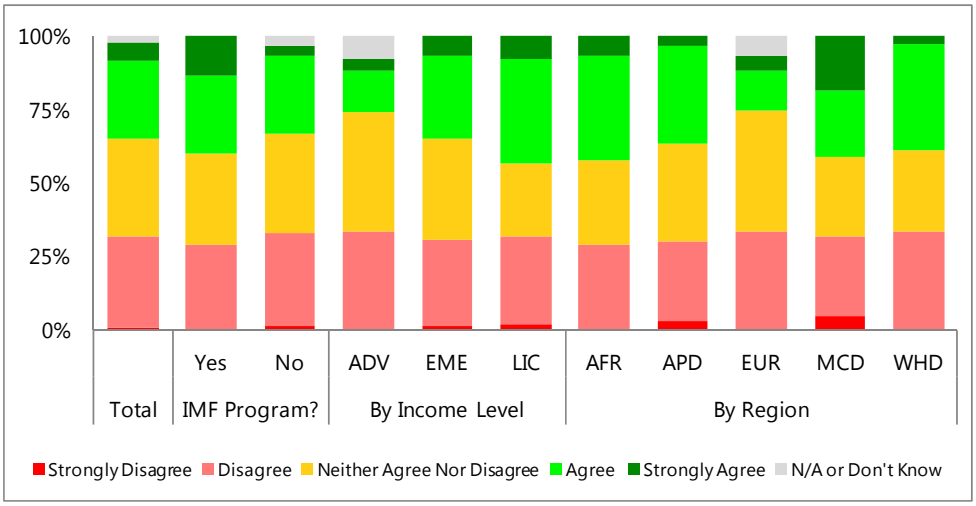
Responding to prompt (b), "For policy discussions, medium-term forecasts are more valuable than one or two-year forecasts," country authorities seem to appreciate medium-term

[2] Among country authorities, but not desk economists, the average response score for program countries is statistically higher (that is, confidence in the medium-term forecasts' underlying assumptions is stronger) than that for non-program countries at the 10 percent significance level (Table A4.2). Table A4.3 shows that economists working on emerging-market economies (EME) are statistically more likely to agree with the statement in (a) than those in low-income economies (LIC). Pairwise *t*-tests (Table A4.5(b)) suggest that desk economists from the Western Hemisphere Department (WHD) are less likely to agree with prompt (a) than those from other departments.

forecasts more than desk economists. The share of the latter who consider medium-term more valuable than short-term forecasts is substantially smaller than the comparable share of authorities. This finding holds regardless of the way the data are sliced, as can be observed by comparing desk economists and authorities' scores in Tables A4.2–A4.4.[3]

When considering different partitions of the data, some statistically significant differences arise among country authorities. Figure A4.1 shows the authorities' responses disaggregated by IMF program participation, country development (income) level, and region. Authorities from countries (i) currently in an IMF program (Table A4.2) and (ii) in lower stages of development, either EME or LIC (Table A4.3), have a more positive view of medium-term forecasts relative to short-term forecasts.[4] Table A4.3 suggests a possible inverse relationship between country level of development and authorities' positive views on medium-term forecasts, but *t*-tests only show statistically significant differences in the responses from authorities in advanced economies—who are less appreciative of medium-term forecasts—relative to both emerging market economies and LICs. No differences exist between emerging market economies and LICs. It is also possible that the less favorable views in advanced economies are a consequence of the current economic situation in some of those economies, more affected by the 2009 financial crisis.

Figure A4.1. Authorities' answers to prompt (b)



Source: IEO Forecast evaluation survey.

[3] The only exception is WHD, in Table A4.3, where the desk economists' views are as neutral as the authorities'.

[4] Authorities in commodity-exporting economies, too, show a propensity to agree that medium-term forecasts are more valuable than short-term forecasts.

In terms of area departments, while *F*-tests detect no statistical difference in response scores (Table A4.4), pairwise *t*-tests in Table A4.5(a) suggest that medium-term forecasts are more highly valued by authorities in AFR countries than in countries covered by the European Department. Given the possible correlation between area departments and countries' level of development, this finding seems consistent with that in the previous paragraph.

Among the responses of desk economists, the inverse relationship between country level of development and more positive views regarding medium-term forecasts is also observed, but is not statistically significant (Table A4.3). The only statistically significant difference is found between the lowest score (i.e., more negative views on medium-term forecasts) from staff in EUR relative to the highest score from staff in the African Department, mimicking the relative value of medium-term forecasts to authorities in their countries of assignment. The fact that there are almost no statistically different scores within the staff responses across different partitions of the survey data, and that the staff scores are systematically lower than both the neutral level (i.e., = 3) and the score from authorities, indicates that a low appreciation for medium-term forecasts is widespread among desk economists at the Fund.

3. Desk economists, the private sector, and authorities do not recommend that the IMF places less emphasis on one- and two-year forecasts and more emphasis on medium-term forecasts.

The responses to prompt (c), that "In its analysis, the IMF should place less emphasis on one- and two-year forecasts and more emphasis on medium-term forecasts," are consistent with those to the previous prompt. Broadly, country authorities, but especially desk economists, do not favor a shift in relative focus from short- to medium-term forecasts by the IMF. Responses from authorities and desk economists are both skewed to the right, concentrated in low scores, and tend to disagree with prompt (c) (see Figure 10 in the main text and Table A4.1).

Although the overall response scores to prompt (c), from both groups of respondents, are lower than for prompt (b) (see Tables A4.1 and A4.2–4), the score from desk economists is statistically significant and notably lower than that from the authorities (Table A4.1). This finding, too, suggests that authorities have a less negative view of medium-term forecasts than staff. It holds regardless of the partition of the data between program and non-program countries (Table A4.2), level of development (Table A4.3), and area department (Table A4.4).

Moreover, also in line with previous findings, the authorities' less negative views on medium-term forecasts are especially true for program countries (Table A4.2) and countries not classified as advanced economies (Table A4.3).[5] Similarly to the case of prompt (b), a

[5] Authorities in commodity exporter countries also seem more favorable to the idea of more emphasis on medium-term relative to short-term forecasts, consistently with their responses to prompt (b).

seemingly inverse relationship between the level of development and positive views on medium-term forecasts (higher scores) is observed in the authorities' responses to prompt (c), although it is not statistically significant. Again, a statistically significant difference only arises between authorities in advanced economies and those in both LICs and emerging market economies. Among desk economists, this relationship is also observed and not statistically significant.

4. Both authorities and staff are relatively neutral on whether medium-term forecasts are too uncertain to be valuable for policy discussions.

Responses to prompt (d) from both groups of respondents, "Medium-term forecasts entail too much uncertainty to be valuable for policy discussions" are very similar and heavily concentrated in the middle, neutral category—which explains the low degree of skewness in both cases and the lack of statistically significant differences between staff and authorities' response scores (Table A4.1).

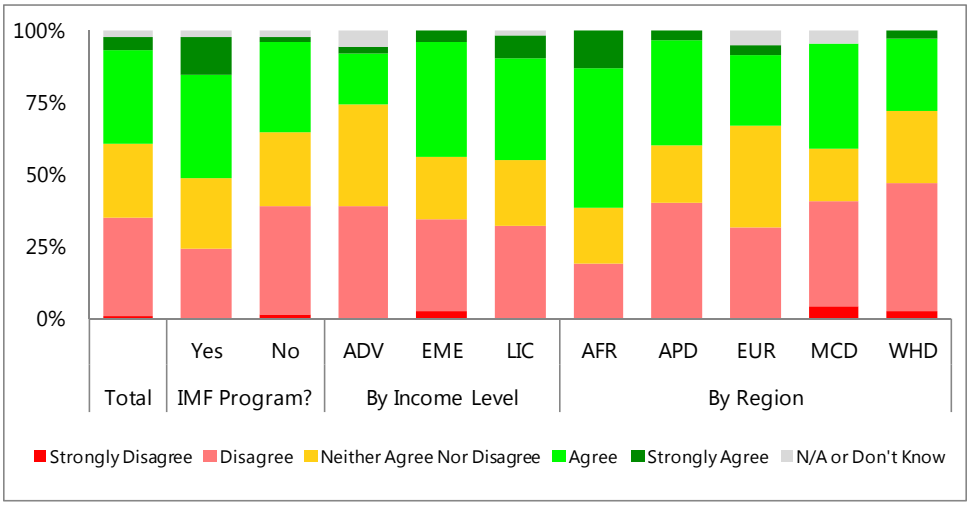
Roughly one-third of respondents disagree and another third agree with prompt (d), with a substantial mass in between these two options (about one-third and one-fourth, among desk economists and authorities, respectively), and very few responses on either edge of the distributions. This pattern suggests that despite the somewhat negative perceptions of both groups with regard to medium-term forecasts *relative* to short-term forecasts, as evidenced in responses to prompts (b) and (c), their *absolute* perceptions on the usefulness of medium-term forecasts are not as negative when assessed from their responses to prompt (d).

Patterns emerge when one looks at different partitions of the data. First, authorities from countries that have recently been in an IMF program are more concerned than authorities from non-program countries that the underlying volatility of medium-term forecasts undermines the use of these forecasts in policy discussions (Figure A4.2 and Table A4.2). This difference is statistically significant. No comparable difference in views between program and non-program countries is seen among staff. Moreover, within program countries, the score from desk economists is substantially lower than the score from the authorities, indicating that the former are less concerned about uncertainty than are the latter.

The more positive views of medium-term forecasts among authorities in program relative to non-program countries—and also in comparison with desk economists, based on responses to (b) and (c)—combined with the stronger reservations shown by authorities from program countries regarding the uncertainty around these forecasts, may seem contradictory at first glance. But it may simply indicate that *because* authorities in program countries value medium-term forecasts relatively more, they are more concerned about their accuracy. Note that, like those for prompts (b) and (c), the response scores for prompt (d) are inversely related to the level of development (Table A4.3), suggesting that authorities from less developed economies, both low-income (LIC) and emerging-market economies (EME), which tend to value medium-term forecasts more highly, also tend to be more concerned

about the uncertainty embedded in these forecasts. These findings are not seen within the responses of desk economists.

Figure A4.2. Authorities' answers to prompt (d)



Source: IEO Forecast evaluation survey.

Table A4.3 also reveals some paradoxical views from desk economists and authorities on the uncertainty of medium-term forecasts. While authorities in advanced economies are the least concerned about uncertainty, among desk economists the opposite is the case, with a substantial difference in scores from advanced economies across the two groups of respondents. Differences in views, albeit in the opposite direction, are also evident among emerging market economies, where desk economists are less concerned than the authorities. Only within low-income economies is there no significant difference between the two groups of survey respondents regarding prompt (d).

Regarding country groupings along the geographical dimension, the underlying survey data in Table A4.4 and Tables A4.5 (a)–(b) confirm the notion that authorities who place a relative larger value on medium-term forecasts, such as those from countries in AFR, tend to be more concerned about the uncertainty in medium-term forecasts, when controlling for certain country attributes such as level of development and IMF program status. Both *F*-tests and pairwise *t*-tests reject the hypothesis of equal average responses from authorities in different area departments. The latter show that scores in AFR are statistically and significantly larger than those in other area departments at the 5 percent level.[6]

[6] By this metric, authorities in oil exporting countries tend to value medium-term forecasts more than desk economists do. Interestingly, this does not hold for commodity exporters, whose authorities tend to agree that medium-term forecasts are too volatile to be valuable, while desk economists working on those countries tend to (slightly) disagree.

A possible implication of the findings regarding prompts (b)–(d) is that desk economists should pay more attention to medium-term forecasts, especially those for program countries and less developed economies.

5. Both staff and country authorities seem comfortable with the assumption that output gaps close within five years.

Despite the potential problems derived from of this assumption, discussed in Section V of the main text and in Annex 2, country authorities and desk economists consider it appropriate for forecasters to assume that output returns to its potential level within five years. A slight inclination to agree with the assumption that "When making medium-term forecasts, it is reasonable to assume that my country's output gap is closed by the end of five years" is reflected in the mild (left) skewness observed in both distributions (Table A4.1). This finding may be explained both by technical difficulties in the desk economists' work and by a perceived lack of realism of the five-year rule in light of the recent financial crisis.

Agreement with prompt (e) is more likely among desk economists and country authorities in program countries relative to non-program countries (Table A4.3), although the difference is not statistically significant. Interestingly, authorities in advanced economies are significantly more supportive of the five-year closing rule than are those in emerging-market economies, but the exact opposite pattern is seen among desk economists: while the authorities in advanced economies are the most likely to agree, desk economists working on advanced economies are the least likely to agree with it. Recall from Section V that the evidence of distortions caused by the five-year-rule is stronger for G7 economies.

Both groups of survey respondents are more in agreement within LICs. In this country group there is no statistical difference in the responses from authorities or desk economists across IMF area departments according to *F*-tests (Table A4.4), although *t*-tests (Table A4.5(b)) suggest that staff in AFR are less likely, and those in MCD more likely, to agree with prompt (e).

Among the authorities from oil-exporting economies, the share of responses agreeing or strongly agreeing with the "five-year rule" for output gaps is much lower—only 8 percent—than in other data partitions, including their counterparts among desk economists (33 percent of whom agree or strongly agree).[7]

[7] Survey results related to oil-exporting economies are available from the author upon request.

6. Desk economists and, especially, the authorities, do not think that when making medium-term forecasts it is reasonable to assume that the real exchange rate (RER) remains fixed for five years.

Both groups of respondents mostly disagree with prompt (f), the assumption that the RER remains fixed for five years. Although the distribution of the authorities' response is mildly skewed to the right, while that of desk economists is essentially symmetrical—which could indicate that the latter are less negative about this assumption than the former—the average response scores are almost the same, and not statistically different (Table A4.1). This finding holds when the survey data are sliced along the program-versus-non-program country dimension; within each of these subsamples, desk economists and authorities' responses are very close.

Looking at the responses to prompt (f) by country development stage, only the desk economists working in LIC economies seem to be neutral, while the those working in both advanced economies and emerging market economies are more negative. Differences relative to the staff in advanced economies, who most clearly disagree with the assumption of a five-year fixed RER, are statistically significant at the 10 percent level, while differences relative to emerging market economies are not significant at that level, but only marginally (Table A4.3).

The results of the *F*-test in Table A4.4 show that among country authorities, the average response scores to prompt (f) differ by area department. The *t*-tests in Table A4.5(a) clearly indicate that the statistically significant differences come mainly from authorities in MCD countries, who are the most comfortable with the assumption.

Among responses from desk economists in different area departments, no statistically significant difference seems to exist, although desk economists in APD, WHD, and EUR have a tendency to disagree with prompt (f), while those in AFR and MCD tend to agree or be neutral about it, respectively. Differences between desk economists and authorities' scores are more pronounced in AFR and MCD.[8]

7. Despite some concerns, as discussed above, country authorities clearly do not ignore medium-term forecasts.

Prompt (g), "We largely ignore the medium-term forecasts," was presented only to authorities and the private sector. Respondents in both groups overwhelmingly disagree (51 percent) or strongly disagree (13 percent) with the statement, while about one-quarter are neutral. Only 10 percent agree (1 percent strongly agree) with prompt (g).

[8] Interestingly, desk economists working on commodity and oil exporter countries tend to dislike the assumption more than those working on non-oil/commodity exporter countries, but authorities in oil exporters largely support the assumption.

Table A4.3 shows that authorities in program countries and less developed economies (both emerging market economies and LICs), respectively, are statistically less likely to ignore medium-term forecasts than are authorities in non-program countries and advanced economies.

Table A4.4 shows substantial differences in response scores from authorities across different area departments. The *F*-test rejects the hypothesis of equal responses across all departments at the 5 percent level. Measured by their responses to prompt (g), authorities in APD countries are the least supportive of medium-term forecasts and are more likely to largely ignore these forecasts, while those in MCD countries are the most supportive.

The disregard for forecasts by authorities from countries in Asia and the Pacific may be related to the fact that the IMF's forecasts for the region are clearly dominated by the forecasts from the private sector. Pairwise *t*-tests in Table A4.5(a) confirm that MCD scores are statistically lower than APD and EUR scores and that the APD scores are statistically higher than those of the other departments, except EUR.

Overview of responses provided by desk economists

Like the responses of country authorities (Table A4.6(b)), responses of desk economists to the survey are internally consistent, as shown by the correlations in Table A4.6(a). Desk economists with more positive views of medium-term forecasts, based on their support for the underlying assumptions about the global economy used in these forecasts (prompt (a)), are also less concerned that medium-term forecasts are too uncertain for use in policy discussions (prompt (d)) and are more likely to support the assumption that output gaps close within five years. Also, those who are more favorable to medium-term forecasts relative to short-term forecasts, according to answers to prompt (b), are more likely favor a shift in emphasis from short-term to medium-term forecasts (prompt (c)) and less likely to be much concerned about uncertainty for policy discussions (prompt (d)). Finally, desk economists who support the assumption that output gaps close within five years also tend to support the assumption that the exchange rate remains fixed for five years.

Table A4.6(a). Correlation across IMF desk economists response scores to prompts (a)-(f)

Correlation *p*-value	(a)	(b)	(c)	(d)	(e)	(f)
(a)	1.00					

(b)	0.29	1.00				
	0.00	---				
(c)	0.05	0.52	1.00			
	0.58	0.00	---			
(d)	-0.19	-0.35	-0.13	1.00		
	0.03	0.00	0.13	---		
(e)	0.16	-0.12	-0.17	-0.09	1.00	
	0.07	0.19	0.06	0.29	---	
(f)	-0.04	0.06	-0.02	0.03	0.24	1.00
	0.64	0.52	0.84	0.70	0.01	---

Note: All correlations and p-values computed using all 126 observations.

Source: Author's calculations using IEO Forecast evaluation survey.

Table A4.6(b). Correlation across country authorities' response scores to prompts (a)-(g)

Correlation *p*-value Observations	(a)	(b)	(c)	(d)	(e)	(f)	(g)
(a)	1.00						

	166						
(b)	0.19	1.00					
	0.02	---					
	166	175					
(c)	0.07	0.45	1.00				
	0.35	0.00	---				
	164	173	174				
(d)	-0.12	-0.30	-0.24	1.00			
	0.11	0.00	0.00	---			
	165	174	174	175			
(e)	0.04	-0.14	-0.02	0.02	1.00		
	0.59	0.08	0.84	0.81	---		
	155	163	163	164	164		
(f)	0.07	0.11	-0.10	0.09	0.07	1.00	
	0.38	0.15	0.22	0.25	0.38	---	
	157	165	164	165	161	166	
(g)	-0.17	-0.31	-0.10	0.23	0.04	-0.11	1.00
	0.03	0.00	0.19	0.00	0.57	0.15	---
	166	175	174	175	164	166	176

Note: The number of observations differs across cases due to the exclusion of responses "not applicable."

Source: Author's calculations using IEO Forecast evaluation survey.

ANNEX 5. IMF PRODUCTS CONTAINING MEDIUM-TERM ANALYSIS WITH RISK ASSESSMENTS AND SCENARIOS

Several IMF products potentially contain discussions about future developments over horizons longer than a year. This annex catalogs the analysis of longer-term issues and scenarios in the flagship publications *World Economic Outlook* (*WEO*), *Global Financial Stability Report* (*GFSR*), and *Fiscal Monitor* (FM), as well as spillover reports, external sector reports, early warning exercises, and Article IV reports.

The *World Economic Outlook* (*WEO)* "presents IMF staff economists' analyses of global economic developments during the near and medium term [which are] integral elements of the IMF's surveillance (…)." The medium-term issues referred to—such as the evolution of commodity prices, effects of structural reforms, fiscal projections, and sustainable economic growth—are an important part of the analysis and are sometimes discussed in special chapters. Of the 28 editions of the *WEO* published since the year 2000, 64 percent have sections or chapters with titles containing key-phrases that suggest medium/long-term subjects;[1] almost 60 percent use of fan charts to describe the uncertainty around baseline scenarios, and more than 90 percent contain figures, charts, or tables to describe medium-term scenarios. The use of fan charts to describe risks to the baseline scenario has been increasing since 2006.

The *Global Financial Stability Report* (*GFSR)* provides a quarterly assessment of global financial markets, with a view towards emerging market financing in a global context. The Report focuses on current conditions and contemporary issues, but discusses financial imbalances and structural deficiencies, extending to the medium term, that could pose risks to financial stability and to emerging-market countries' ability to borrow. In particular, the Report draws out the financial implications of economic imbalances highlighted in the *WEO*, including on medium-term projections, to assess current risks. The *GFSR* often also contains articles and analytical chapters on structural or systemic issues that deal with medium-term analysis.

The *Fiscal Monitor (FM)* began to be produced in response to the fiscal challenges in the aftermath of the 2007–09 financial crisis. It focuses on the multilateral surveillance of fiscal issues, aiming at analyzing the latest public finance developments, updating fiscal implications of the crisis and medium-term fiscal projections, and assessing policies to put public finances on a sustainable footing. The FM's projections—including the medium-term fiscal projections that incorporate policy measures judged by the IMF staff as likely to be implemented, IMF program projections, and estimates of cyclically adjusted primary balances—are based on the same database used for the *WEO* and *GFSR*.

[1] The key-phrases searched by the evaluation team, which by no means exhaust all possibilities, are "medium term, long term, potential output, output gap, sustainable growth, structural reform, structural change, debt sustainability, demographic change."

Compared to the *WEO*, the discussion of medium-term issues is more frequent in the *Fiscal Monitor*[2] where it is mostly restricted to fiscal issues, although the use of fan charts and tables/figures for medium-term scenarios is substantially less frequent than in the *WEO*. Among the flagship products, the *GFSR* seems to be the least concerned with medium-term issues and scenarios. However, even if medium-term issues are not the central part in the *FM* and *GFSR*, they often feature in the overall analysis and discussion. For instance, a simple inspection of the contents of three recent editions of both the *FM* and *GFSR*[3] reveals that between 60 percent and 70 percent of all sections deal, at least partially, with structural issues and medium- to long-term trends, prospects, or risks.

Article IV reports use charts and tables containing medium-term scenarios more frequently than the *WEO*, although their use of fan charts is relatively rare (only 4 percent). Within a random sample of 56 Article IV reports selected among the most recently available for the entire IMF membership, 96 percent use such charts and tables and the number of such tools per report is more than twice that in the *WEO*. Moreover, 84 percent of Article IV reports in the sample contain sections or chapters with titles suggestive of longer-term subjects.

Spillover Reports, Pilot External Sector Report, and Early Warning Exercises

Spillover reports examine the external effects of domestic policies in five systemic economies (S5): China, the Euro Area, Japan, the United Kingdom, and the United States. The reports complement the Fund's Article IV discussions with these economies and serve as inputs into the Fund's multilateral surveillance, by analyzing the transmission channels of monetary, exchange rate, fiscal, financial, and structural policies between the S5 economies and the global economy with a view to anticipating the cross-border impact of policies. The identification and assessment of policy spillovers with the greatest potential impact—based on staff macro models, inputs from a multi-departmental team, and individual discussions with authorities—also aims at complementing the discussion of the risks to global economic and financial stability in the *WEO* and *GFSR*, and points to possible areas for policy coordination.

Also discussed in spillover reports are medium-term issues such as the potential impact of financial sector and structural reforms affecting potential growth, long-term fiscal measures, and monetary/exchange rate policies or frameworks. Table A5.1 lists all specific examples of discussions involving medium-term projections in the three issues of the spillover report.

---

[2] Table 1, in the main text, shows that when considering the period 2006–2013, which better captures the period in which the *FM* has been available (since 2009), the frequency of reports with sections about medium-term issues is 60 percent in the *FM* (1.4 sections about the subject per report) compared with 50 percent in the *WEO* (0.9 sections per report).

[3] The three editions are for Spring and Fall of 2012 and Spring of 2013.

Note that the number of explicit quantitative analyses of developments over the medium term in the reports has risen every year.

Table A5.1. Medium-term analysis in spillover reports

Vintage	Analysis
2011	• The cumulative effects of planned fiscal adjustments in major economies (US, Euro Area, and Japan) on the output of other S5 economies over next five years. • The medium-term impact of an appreciation of the Chinese exchange rate on the GDP of other economies.
2012	• The impact of the Euro Crisis on the fiscal space and reserve coverage needed to reach a desired target for the debt-to-GDP ratio in low-income countries. • The effect of a credible medium-term fiscal adjustment in the US on the erosion in public confidence. • The cumulative response of oil prices to shocks to global liquidity over next 20 quarters. • The effect of rebalancing the investment and consumption ratios in China on the GDP of other economies and on commodity prices. • Effects of an increase in Japanese bond yields on global yields and economic growth.
2013	• The long-run effect on the GDP of policies in S5 economies leading to a reduction in the risk of adverse spillovers to global economy. • Scenarios about the cumulative effect of QE announcements on GDP. • Structural reforms to increase potential output in the EU Area and Japan, to reduce risks during the transition to a higher consumption-to-GDP ratio in China, and structural fiscal reform in the US and Japan. • Positive net growth spillover effects from new monetary policy stance ("Abenomics") over the long run. • Cumulative impact over 3 years of a smooth normalization of monetary policy in the US on GDP. • The effect of "rebalancing policies"—needed to reduce the imbalances of global current accounts and in domestic policies in S5 economies—on global GDP over 10 years.

The *Early Warning Exercise* (EWE)—which the Fund conducts jointly with the Financial Stability Board (FSB)—emerged from the need to improve the ability of multilateral surveillance to flag risks and vulnerabilities that could come from systemic shocks, such as those leading to the recent financial crisis. The EWE focuses on low-probability but high-impact risks to the global economy and on policies to mitigate these risks, integrating macroeconomic and financial analyses and using a number of quantitative tools and broad-based consultations. No report is made available to the public; findings are confidentially presented to senior officials during the IMF Spring and Annual Meetings. The EWE typically contains less medium-term analysis than spillover reports, although the initial EWE rounds concentrated primarily on potential mutations of the 2007–09 financial crisis, asking what new shocks could materialize and assessing the consequences of policy inaction over an unspecified horizon that may include the medium term. It is expected that once the global economy returns to more stable conditions, the EWE will become more forward-looking as initially planned.

The Pilot *External Sector Report* (ESR) provides a snapshot of multilaterally consistent analysis of the external positions of 28 large economies and the Euro Area. It combines insights from IMF staff on individual economies with multilateral analysis about exchange rates, current accounts, balance sheet positions, reserves adequacy, and capital flows. One premise is that current account imbalances and deviations of exchange rates from a desired "norm" may be useful for the assessment of member countries' overall economic and financial policies, to the extent that those gaps reflect the joint effects of policies targeted both at the domestic economy and the external sector as well as of structural factors (e.g., demographics, social protection schemes etc.).

Medium-term or structural issues have been discussed in the two issues of the ESR; a list of all specific topics is provided in Table A5.2.

Table A5.2. Medium-term analysis in external sector reports

Vintage	Analysis
2012	• Moving current accounts toward fundamentals likely implies ambitious medium-term policies and significant real exchange rate realignments. • Adjustments to structural factors are needed to reduce vulnerabilities to external imbalances (e.g., changes in social protection frameworks that affect precautionary savings). • Expected medium-term policy changes (as announced and discussed in the most recent *WEO* before publication of the 2012 ESR) are likely to produce only modest effects on the current account divergences over the next five years. • Differences in cyclically adjusted current account balances and current account balances consistent with fundamentals and desired policies are used as a measure of undesirable external imbalances. The estimation of cyclically adjusted variables requires estimates of the long-run sustainable output level (potential output).
2013	• Medium-term policies to close structural policy gaps and reduce undesired current account imbalances include fiscal consolidation over the medium term and structural reforms in deficit countries. • Discussion of risks of prolonged use of extraordinarily low interest rates and quantitative easing in the US. • "Looking ahead" section analyzes past data on the determinants of capital inflows, distinguishing between structural and temporary factors. Although the horizon is not specified the forward-looking analysis presumably involves longer horizons.

The analysis in the ESR, which encompasses the Pilot External Balance Assessment (EBA), mostly concentrates on the short-term assessment of policies rather than on projections of future outcomes. Nevertheless, assumptions about a country's long-run sustainable output level and growth rate, sometimes embedded in point forecasts of GDP growth for the medium term, are required by the EBA approach, which is combined with judgment to help assess external imbalances, as discussed in Section II of the main text.