IEO
Independent Evaluation Office
of the International Monetary Fund

BACKGROUND PAPER

BP/14/04

On the Accuracy and Efficiency of IMF Forecasts: A Survey and Some Extensions

Hans Genberg and Andrew Martinez

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/14/04

IEO Background Paper

Independent Evaluation Office
of the International Monetary Fund

On the Accuracy and Efficiency of IMF Forecasts: A Survey and Some Extensions

Prepared by Hans Genberg and Andrew Martinez

February 12, 2014

Abstract

This paper presents new perspectives on IMF forecast accuracy and efficiency by combining established analytical approaches with up-to-date data on forecasts and outturns and by developing new methodologies to help draw practical lessons for IMF forecasting. We find that the IMF forecasts developed for the *World Economic Outlook* are not consistently biased in one direction or the other, nor do they consistently perform better or worse than those of comparators. Even so, they have tended to be consistently over-optimistic in times of country-specific, regional, and global recessions. Second, while the Fund's forecasts incorporate interdependence among economies to a significant degree, the data show that they may need to take better account of the international repercussions of developments in the Chinese, German, and US economies. Third, experience matters: staff with longer experience, whether country-specific or general, make smaller forecast errors, especially in the case of low-income countries.

JEL Classification Numbers: C530

Keywords: Forecasting, IMF, *World Economic Outlook*

Authors' E-Mail Addresses: Genberg@verbier.ch, andrew.martinez@economics.ox.ac.uk

Contents Page

ABBREVIATIONS

EC	European Commission
EIU	Economist Intelligence Unit
EPA	earliest published actual
EPE	earliest published estimate
FAD	Fiscal Affairs Department
G-7	Canada, France, Germany, Italy, Japan, United Kingdom, United States
GDDS	General Data Dissemination Standard
HRD	Human Resources Department
MAE	mean absolute error
OBP	Office of Budget and Planning
OECD	Organization for Economic Cooperation and Development
RMSE	root mean square error
RMSFE	root mean square forecast error
SDDS	Special Data Dissemination Standard
WEO	World Economic Outlook

I. INTRODUCTION

1. This paper assesses the accuracy and efficiency of IMF forecasts by drawing on empirical results found in the literature and by providing new evidence bearing on how to interpret these results. We focus throughout on forecasts published in the *WEO*, as these are the most easily accessible and by far the most frequently analyzed in the literature. For the same reason most of our discussion refers to forecasts for GDP growth for the current year and for one-year ahead, as these have been the variables subject to most analysis in the literature.

2. We start by describing the form of the tests most often used for assessing the accuracy and efficiency of forecasts and point to reasons why caution needs to be applied in interpreting the results of these tests. We use up-to-date information to interpret sometimes contradictory evidence found in the literature.

3. In Chapter III of the paper we present new evidence. In Section III.A, we build on studies that have found that private sector forecasters are particularly prone to make errors in forecasting recessions. We show that *World Economic Outlook (WEO)* forecasts are no better in this respect, and then use these results to suggest that the "optimistic" bias in *WEO* forecasts may be a manifestation of difficulties in predicting recessions. In Section III.B, we show empirically that the experience of the forecaster—the country desk economist in the case of the IMF—makes a difference for the accuracy of the final forecast. For the IMF, this has implications for training, for the tenure of staff at a country desk, and for the process by which information is passed from the incumbent to the successor in a country assignment. New evidence on the efficiency of *WEO* forecasts, specifically on whether these incorporate interrelationships between economies, is presented in Section III.C. The results indicate that while *WEO* forecasts do incorporate interdependencie*s* among economies to a substantial degree, some linkages still seem to be unaccounted for.

4. Chapter IV concludes, drawing three key lessons from the analysis.

II. THE QUALITY OF IMF FORECASTS: A VIEW FROM THE LITERATURE

5. To put the literature on the accuracy and efficiency of IMF forecasts into perspective, we first present the criteria typically used in assessing forecast quality. We then briefly review the findings in the published empirical literature. We end this chapter by pointing to some caveats that need to be kept in mind when interpreting the results from typical tests of accuracy and efficiency. We illustrate the possible consequences of these caveats by presenting results based on standard empirical techniques applied to the most up-to-date data.

A. Criteria for Evaluating Forecast Performance

6. There exists an extensive literature on forecast evaluation.[1] The purpose of this section is simply to highlight some features that have received particular attention in the context of evaluations of IMF forecasts.

(i) Bias

7. The most common concern expressed about IMF forecasts is that they are biased. The Oxford English Dictionary contains two definitions of the term biased.[2] One is associated with statistics: "Containing a bias or error which will not balance itself out on average." This is what we have in mind here. Some of the academic literature and some popular discussions of IMF forecasts use another definition that is more tendentious, namely: "Influenced; inclined in some direction; unduly or unfairly influenced; prejudiced." In Section II.B (ii) below, we review briefly the literature that appears to take this definition as a point of departure.

8. The statistical definition of bias is simply the average forecast error in a sample of forecasts. Specifically, let $FE_{t,t+h}$ stand for the error of the forecast (denoted $\hat{y}$) at time t of a variable y at time $t+h$ (h for horizon):

$$FE_{t,t+h} = y_{t+h} - \hat{y}_{t,t+h} \qquad (1)$$

then the statistical bias B (in a sample S of h-period forecast of variable y) is simply

$$B(S,h,y) = \sum_{t}^{t+s-1} \frac{1}{s} FE_{t,t+h} \qquad (2)$$

where s is the sample size.

9. Tests of significance could be carried out either based on B or, equivalently, by a regression of the forecast error on a constant. A slightly more general test would be to estimate the so-called Mincer-Zarnowitz regression in which the future actual value is explained by a constant and the forecast.

(ii) Efficiency

10. The efficiency of forecasts refers to whether or not they take into account "all available information." Since "all available information" is a large set, it is much more difficult to reject inefficiency of a forecast than it is to reject biasedness.

---

[1] See, for example, Diebold and Lopez (1996) and West (2006).

[2] See www.oed.com/view/Entry/18566 .

11. Regression-based tests for efficiency can be carried out by regressing the forecast errors on variables whose values were known at the time the forecasts were made. A particularly common test is to regress the current forecast error on past periods' forecast errors. Efficiency would, in a stationary environment, imply the absence of serial correlation in the errors, which could be tested by such a regression. Faust (2013) however argues that in a context where structural change is ongoing, it is possible that serial correlation of forecast errors simply reflects learning on the part of the forecaster about the economy and would therefore not be a sign of inefficiency.

12. In the context of the IMF's multi-country forecasts, a particularly interesting question relating to efficiency is whether the forecasters in each individual country take proper account of interdependencies among all the Fund member countries. Timmermann (2006) tested for this by asking whether forecast errors were in part explainable by the forecasts of the U.S. and German economies—forecasts that were already known when the forecasts for other countries were made. He showed that this was indeed the case. In Chapter III, Section III.C, we follow this line of inquiry using an extended sample period. We also develop and apply a methodology to test whether IMF forecasts incorporate global determinants of domestic GDP growth to the same extent as is found in the data.

(iii) Comparison with other forecasters

13. A significant fraction of the literature assesses the forecasts of the IMF by comparing them with forecasts of other institutions such as the Organisation for Economic Co-operation and Development (OECD), national central banks or treasuries, and private sector bodies. Such comparisons usually employ measures such as the mean absolute error, mean square error, or the root mean square error, and can provide useful indicators of which forecasts are most accurate at the time they are published.

14. But if the objective is to assess the ability of the forecaster or the soundness of the forecast process, such comparisons can be misleading. The reason is that even if the forecasts by two different institutions are published at the same time, one forecast may be based on more recent information than the other. For example, consensus forecasts[3] published at the same time as the Spring or Fall *WEO* forecasts are almost certainly based on more recent information; it is well known that private sector forecasters are more nimble than the IMF at incorporating new information (see Genberg, Martinez, and Salemi, 2014). The problem of comparing forecasts from different institutions is of course compounded if their publication dates differ substantially.

[3] Consensus forecasts are published by Consensus Economics; they are typically the average of all the private forecasts surveyed by that organization.

B. Results in the Literature

(i) Tests of biases

15. Of papers that test for biases in IMF forecasts a large proportion focus on growth forecasts. Finding an optimistic bias in growth forecasts is a frequent feature of this literature; examples include Aldenhoff (2007), Artis (1988), Hawkins (2002), and Timmermann (2006, 2007). But a handful of papers—Takagi and Kucur (2006, 2008) and Krkoska and Teksoz (2009)—conclude that some IMF growth forecasts are unduly pessimistic, and some—e.g., Beach and others (1999)—report results indicating that IMF forecasts in some cases are unbiased.

16. There are several potential reasons for differences in results among studies. One is the choice of country groupings. For example, GAO (2003) concludes that IMF growth forecasts for countries with IMF supported programs are biased (optimistic) whereas those for non-program countries are not biased;[4] Beach and others (1999) conclude that IMF growth forecasts are pessimistic for developed countries but optimistic for developing countries; and Takagi and Kucur (2006) conclude that IMF forecasts are optimistic for Africa and in IMF program cases but pessimistic for industrial countries. It is noteworthy that the results in the most comprehensive of all studies—Timmermann (2006, 2007)—also contain many nuances indicating that biases often differ in sign between countries and are frequently not statistically significant.

17. Some authors have also noted that biases may vary over time—Boettcher (2004) —or be due to the presence of unanticipated recessions in the sample period—Abreu (2011). We return to these issues later in this chapter and in Section III.A.

18. Tests of possible biases in the Fund's inflation forecasts yield results similar to those found for growth forecasts, namely that the sign and significance of any bias found tends to vary with the country and time period studied. Thus, the early study by Kenen and Schwartz (1986) concludes that forecasts under-predict the subsequent outcomes—a conclusion shared by Timmermann (2006, 2007) though only for African, European, and Western Hemisphere economies. Beach and others (1999) report that the Fund's inflation forecasts for developed countries are unbiased and efficient.

19. The diverse results found in the literature are confirmed in responses to a survey undertaken for this evaluation. Country authorities, central bank and finance ministry officials, and forecasters in the private sector were asked about their perception of *WEO* forecasts.[5]

---

[4] Luna (2014b) contains a more thorough review of the literature focusing on program countries.

[5] Genberg and Martinez (2014) describe the survey in more detail.

20. Asked whether they in general viewed *WEO* forecasts as unbiased, 57 percent of country officials responded that they either "strongly agreed" or "agreed." Nine percent answered either "strongly disagreed" or "disagreed." The remainder of the respondents either had no opinion or "neither agreed nor disagreed." These response patterns were very similar across country groupings by area, income level, and whether or not the countries were being supported by IMF-supported programs.

21. To a more specific question about the accuracy of *WEO* growth forecasts for their own country, 76 percent of country officials responded that they believed these forecasts were "about right." Six percent believed the forecasts were "consistently too high" and 18 percent said they were "consistently too low." For program countries the responses were similar: 76 percent of the respondents said "about right," 9 percent "consistently too high," and 16 percent "consistently too low."

22. Respondents working in global financial institutions had less sanguine views about the accuracy of *WEO* forecasts. Fifty percent believed the forecasts were ''about right," while 27 percent thought they were "consistently too high," and 23 percent "consistently too low."

23. These survey results are interesting because they suggest that country authorities by and large do not question the quality of IMF forecasts. Of course one can argue that when 24 percent of officials feel that *WEO* growth forecasts are consistently either too high or too low, this is already a sign that something is amiss. It is also noteworthy that three-quarters of these officials think that growth forecasts are too pessimistic rather than too optimistic.

(ii) Explaining forecast errors

24. A broad range of studies examines and attempts to explain the IMF's forecast errors. This literature focuses mainly on four areas: program-related concerns, data limitations, initial assumptions, and potential political influence.

25. The most popular strand in this literature examines how issues related to IMF-supported programs may affect forecast errors. Beach and others (1999) found that the amount of IMF lending could explain a significant proportion of the errors in forecasts for developing countries. Some later studies have found that forecast errors in IMF-supported programs are associated with program success and completion (Golosov and King, 2002) as well as with program size (Musso and Philips, 2002 and Luna, 2014b). Others have found that forecast errors in IMF programs are primarily driven by mismeasurement of the initial conditions and model misspecification (Atoyan and others, 2004).

26. Data quality and assumptions affect forecast errors not only in IMF program cases. Tong (2004) and Mrkaic (2010) find that IMF forecast errors are related to whether or not a country subscribes to the IMF's Statistical Data Dissemination Standards (SDDS). Similarly, Takagi and Kucur (2006) find that the IMF's macroeconomic forecast errors are largely

correlated with errors in the underlying assumptions about oil prices and interest rates and argue that better assumptions could improve the IMF forecasts. More recently, Blanchard and Leigh (2013) find that errors in GDP forecasts are explained by forecasts of the fiscal balance and argue that the forecasters must have therefore underestimated the size of the fiscal multipliers.

27. An alternative strand in the literature explores how political influence may affect IMF forecast errors. Aldenhoff (2007) finds that election years in some advanced economies are associated with an optimistic bias in IMF forecasts and argues that the IMF follows a "political forecasting cycle." Dreher and others (2008) test for several different types of political influence on IMF forecasts and find that optimistic forecasts are associated with countries whose votes in the U.N. are similar to those of the U.S.

(iii) Tests of efficiency

28. Timmermann (2006) found that taking account of growth forecasts for Germany and the U.S. could have improved the Fund's forecasts for certain other countries during his sample period (1991–2003). Others have tested efficiency by comparing IMF forecasts with pure time-series models. Arora and Smyth (1990) concluded that for six groups of developing economies a pure random-walk model performed better than *WEO* forecasts. Zeng (2011) found a time-series model with a better forecasting record for GDP growth for Korea in the 2000–08 period, although he also found that a number of other models yielded forecasts inferior to *WEO* forecasts.

29. But here, too, some authors reach different conclusions. We have already noted that Beach and others (1999) conclude that growth forecasts were efficient in developed countries. Similarly, Elliot and others (2005) do not reject the rationality of *WEO* forecasts of budget-deficit-to-GDP ratios for G7 countries. Ashiya (2006) finds differences across countries: IMF forecasts fare well in tests of the rationality of forecast revisions for GDP and the GDP deflator for some countries (Germany, Italy, and the U.K.) but not for others (Canada, France, and Japan).

30. Like the results of tests for biases, the results of tests for efficiency seem to differ according to what country or country groups are studied and according to the sample period.

(iv) Comparing the IMF and other forecasters[6]

31. The IMF's *WEO* forecasts are often viewed as a benchmark when comparing against other national and international forecasts. A survey conducted for the Independent Evaluation Office of the IMF (IEO) found that almost 88 percent of country authorities either agreed or strongly agreed with the statement that they "consider the *WEO's* projections to be the benchmark for assessing economic prospects" (IEO, 2006). More recently, in the survey

[6] See Annex 3 for a more comprehensive discussion of this literature.

conducted for the present evaluation, 64 percent of the country authorities who responded either agreed or strongly agreed with the statement that they "use *WEO* forecasts to check the accuracy of [their] own forecasts" (see Genberg and Martinez, 2014).

32. Since the *WEO* was first published in the 1980s, studies have regularly compared its forecasts with those of other bodies. The comparators have typically included the OECD (see Table 1), but private forecasting aggregators such as Consensus Economics have increasingly featured in the comparisons.

Table 1. Most frequently used IMF comparators in the literature
(In percent of studies)

OECD	62
Consensus Economics	49
European Commission (EC)	31
Econometric models[1]	18
World Bank	11
Economist Intelligence Unit (EIU)	9

Source: Authors' calculations based on 45 studies.
[1]Does not include naïve rules such as 'no-change' forecasts.

33. As noted above, comparing the forecast performance of different institutions should be done cautiously, particularly when their forecast-release dates differ. As shown in Table 2, relative to its main forecast comparators, the *WEO* is released relatively early in each forecasting cycle. This means that the IMF's Fall forecast could be published up to three months before the OECD's forecast—giving the OECD (and other forecasters) time to incorporate the IMF's forecast as well any new information that may emerge in the interim. While these timing differences could have an important impact on relative forecast performance, only a few past studies make more than a passing note of the differences in production dates.

Table 2. Average difference in publication dates with *WEO*, 2000–12[1]
(In number of days)

	OECD	Consensus[2]	EC	World Bank	EIU[2]
Spring/Summer	62	-1	2	57	11
Fall / Winter	87	-2	42	75	24

Source: Authors' calculations.
[1] A positive number indicates that the *WEO* forecasts are published first.
[2] Publication dates for these institutions, whose forecasts are released on a more frequent basis, were chosen to minimize the distance from *WEO* publication dates.

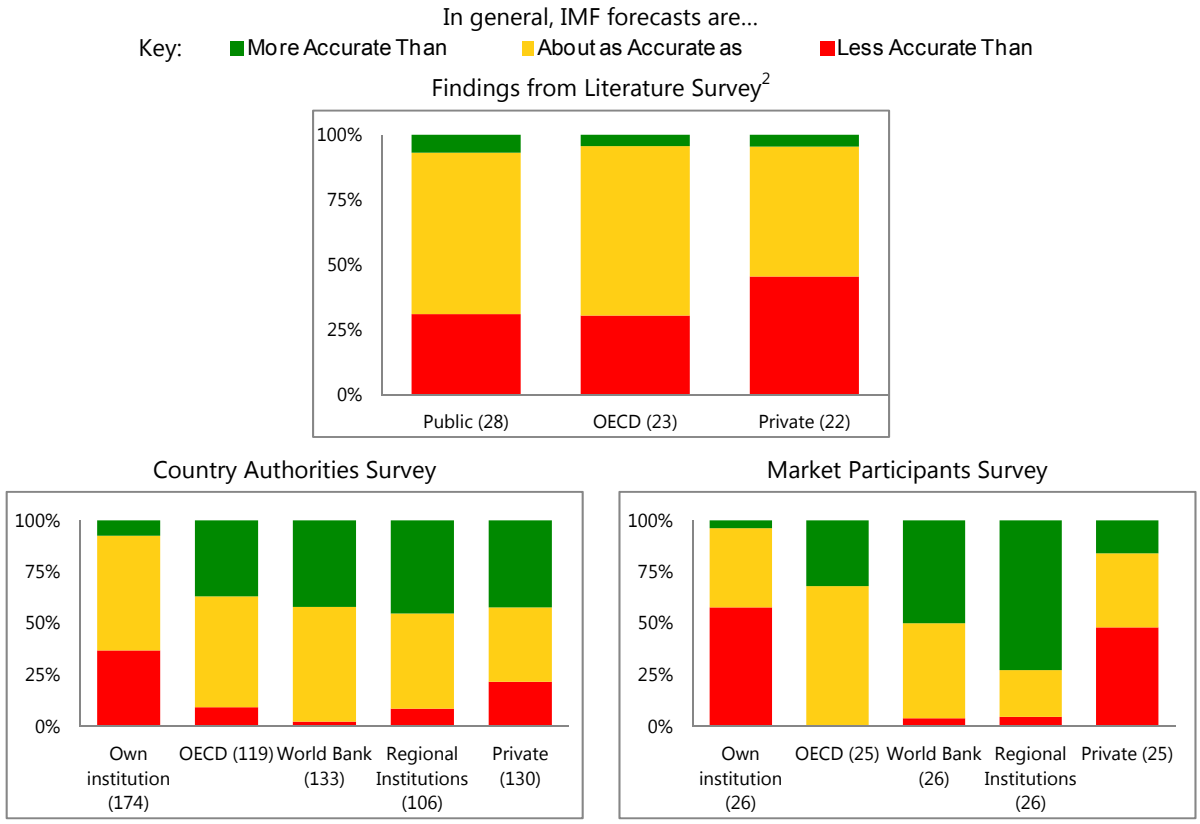
34. Less of a publication timing issue arises when comparing *WEO* forecasts with private forecasts. This is largely because private forecasters for Consensus Economics and the Economist Intelligence Unit issue their forecasts monthly and thus the appropriate publication date can be selected so as to minimize the differences. Even so, while the forecasts from these sources are released without much delay, *WEO* forecasts as produced by the Fund's individual desks are typically one month old by the time they are published (see Genberg, Martinez, and Salemi, 2014). The Fund's lengthy production process may penalize the performance of its forecasts.[7]

35. Overall, the findings of the literature comparing IMF and other forecasters appear to be mixed, with performance depending on the variables examined, the set of countries, and the time period. In general, many studies find that IMF forecasts perform on par with those of other forecasters. Where they find differences, they are more likely to find that the IMF

[7] Timmermann (2006, 2007) finds that the selection of different months of consensus forecasts does affect their relative performance vis-a-vis the *WEO* forecasts.

performs slightly worse than other forecasters, although differences in timing complicate the comparisons (Figure 1, top panel).

Figure 1. Perceptions versus Reality: How does the IMF compare against other forecasters?[1]



Source: Authors' calculations and IEO Forecast Evaluation Survey.

[1] Sample size for each category in brackets.

[2] Based on a review of 45 studies comparing IMF forecasts with other forecasters.

Perceptions of country authorities and market participants

36. The results of the survey undertaken for this evaluation suggest that country authorities and market participants have a higher opinion of IMF forecasts than the literature would suggest (Figure 1, bottom panel).[8] Comparing IMF performance with that of OECD, one-third of the country authorities and market participants surveyed—compared with only 5 percent of the studies reviewed—thought that IMF forecasts were more accurate. And comparing the Fund with private forecasters, 42 percent of country authorities and 16 percent of the market participant survey respondents—compared to only 5 percent of the studies reviewed—thought that IMF forecasts performed better than private forecasts. It is possible

[8] The literature has an overrepresentation of advanced and European economies in its sample relative to the survey population which includes country authority responses from all over the world. However, limiting the survey responses to just those authorities in advanced or European countries does not significantly change the results.

that country authorities incorporate differences in the forecasts (such as timing) when judging relative performance.

37. While country authorities and market participants tend to have a more positive view of the IMF than of other external forecasters, their views are more nuanced relative to their own forecasts. Interestingly, market participants (i.e., private forecasters) have a more positive view of their own forecasts that is broadly in line with that in the literature. Similarly, for their own economy, country authorities have a much more critical view of the IMF's forecasts relative to their own, with authorities from advanced economies somewhat more negative than those from low-income countries. These results appear much more consistent with the results from the literature and with the view expressed by some country authorities that they know their country's economy much better than the IMF does.

C. Accounting for the Diversity of Results

38. In view of the variety of results found in the literature we discuss three factors that may account for the differences across the studies: aggregation; how to measure the actual value against which a forecast is compared; and the importance of the sample period. Then, using standard methods from the literature, we present additional empirical results that provide new perspectives on the quality of IMF forecasts.

(i) Aggregation

39. Tests for biases in IMF forecasts are often carried out on aggregates of countries, the most common being program versus non-program countries, low-income countries, emerging markets, and advanced countries. In doing so it is implicitly assumed that each country group is homogenous, but while this may be true in some dimensions it is not necessarily so in the case of forecasts. For example, emerging markets may have common characteristics that make them grow at a different rate than advanced economies, but this does not necessarily influence the nature of the forecasts for these economies. On the other hand, it can be argued (see, for example, Faust, 2013 and Luna, 2014b) that the nature of the forecast process in the context of program countries is very different from that in the context of nonprogram countries. If so, tests of biases for each group separately are advisable.

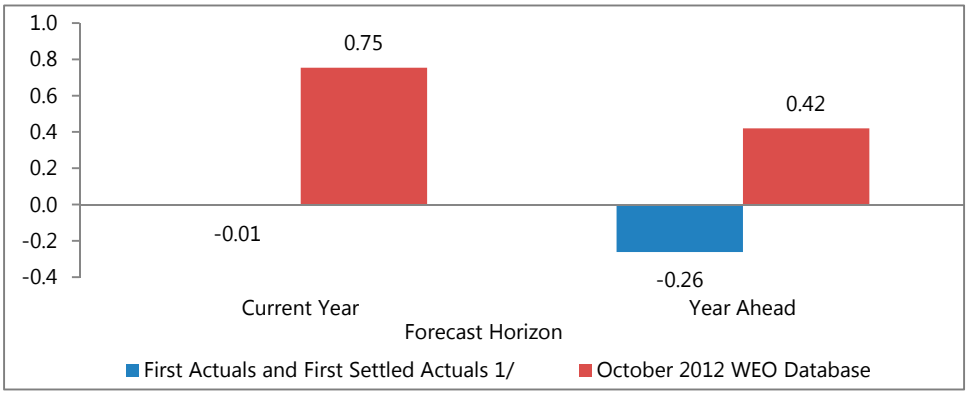
(ii) How to measure actual value?

40. To calculate forecast errors it is obviously necessary to identify the actual value of the variable being forecast. What is less obvious is how to choose this value. This question arises because data revisions can lead to differences between the value ascribed to a variable when it is first published and the value ascribed, for the same time period, some years later after data revisions have been made.

41. In some cases the difference between "earliest available" and "latest available" can be quite stark. Figure 2 illustrates the case of *WEO* forecast errors for GDP growth in the United

Kingdom based on the period 1990–2003; the same data are used in Timmermann (2006, 2007). Using earliest available "actual" values yields a slightly negative bias (over-prediction) in the average current-year forecast and a relatively large negative bias for the average next-year forecast. This is in contrast with the biases computed with the data for the same GDP growth rates that were published in 2012 for the entire period.

Figure 2. Average forecast errors of IMF forecasts of the UK's real GDP, 1981–2011
(In percentage points)



Source: Authors' calculations using IMF, *WEO*.

[1] "First Actuals" are from the Spring *WEO* database of the year after each current year forecast. "First Settled Actuals" come from the Fall *WEO* database of the year after each year ahead forecast.

42. This point can be made more formally by calculating the simple average of Spring current-year forecast errors for GDP growth for all years from 1991 to 2011 and for 139 countries for which we have a complete set of forecasts and actuals.[9] When the "first actual" is used as the actual value in the calculation of the forecast error, the bias turns out to be a (statistically highly significant) value of -0.3, indicating an optimistic bias. When instead the values contained in the 2012 *WEO* data base are used, the bias is 0.02 (it too is highly statistically significant).

43. There seems to be a consensus in the literature that for comparison with forecasts the first available measures of "actual" values of a variable are preferable to the latest available measures.[10] Of the studies we reviewed, about 25 percent used the latest available estimate, about 40 percent used the first available actual, and 6 percent used a settled estimate. The IMF itself appears not to have a view on whether the earliest published estimates or earliest published actual figures should be used; it has often used the latest available estimates, though the recent IMF guidance on debt sustainability analysis used estimates from the Spring t+2 vintage of the *WEO*. Our analysis in this paper is based primarily on the first actual values, but we note where there are substantial differences using settled values.

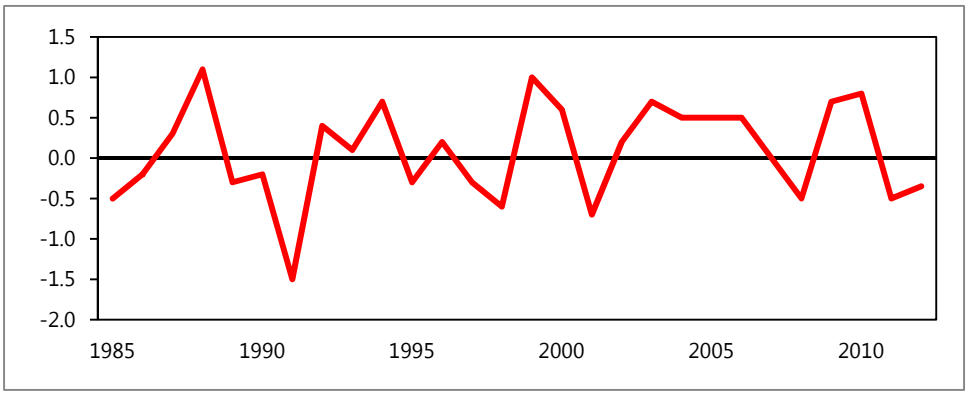
[9] The total number of observations is therefore 2919.

[10] See, for example, Timmermann (2006, 2007) and Artis (1996).

(iii) The importance of sample period

44. Forecast errors are not constant over time.[11] Figure 3 illustrates this fact using the September/October forecasts of next-year world GDP growth rates as published in the *WEO*. Some periods are more tranquil than others and are therefore less likely to cause difficulties for a forecaster. Others are dominated by major shocks—the great recession starting in 2008 being the most recent example—which lead to substantial forecast errors. Given that tests of a forecaster's bias always have to be carried out in relatively small samples, the results may be influenced by special characteristics of the business cycle during the sample period.

Figure 3. World: Forecast error of GDP growth: Spring current year forecasts
(In percentage points)



Source: Authors' calculations using IMF, *WEO*.

45. The issue hinges on whether the stochastic process driving the forecast errors is stationary. In the context of IMF forecasts this may be questioned. As noted in Genberg, Martinez, and Salemi (2014), IMF forecasts are the responsibility of the country desk economists that the IMF assigns to each of its member countries. Two facts about the forecast process used by desk economists are important for the present discussion: the forecasts are heavily influenced by judgment, and economists tend to move between country assignments after at most two to three years. Unless the methods and judgment of successive country desk economists are very similar, it would not be surprising if the properties of the forecasts—and hence of the forecast errors—for a given country were to change over time.

46. How important is the sample period for the major issues regarding IMF forecast quality discussed in the literature? Below we present evidence that suggests that it can be decisive.

[11] If they were, a trivial correction of the forecast would remove the error.

Estimates of biases

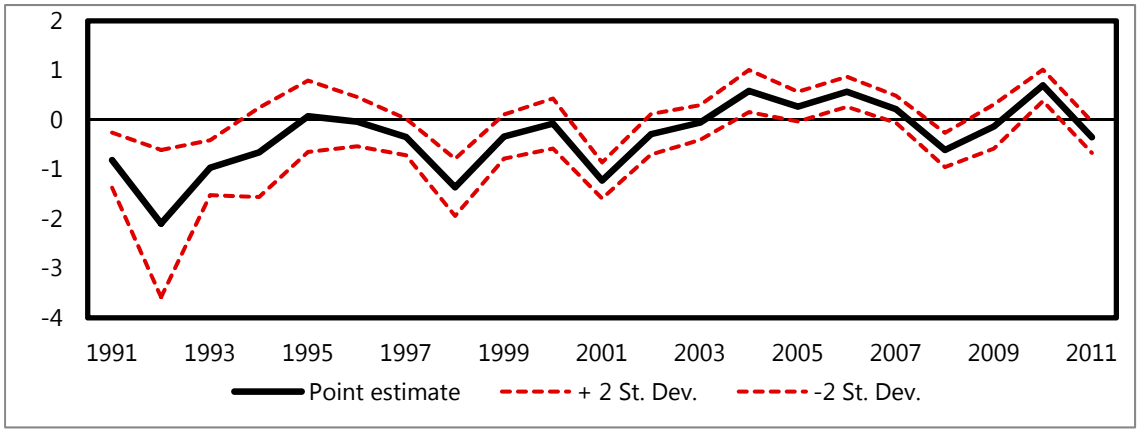
47. To provide an interpretation of sometimes contradictory results in the literature, we test for bias in the sample mentioned earlier of forecasts of GDP growth and inflation for 139 IMF member countries for the period 1991–2011.[12] We focus on current-year forecasts published in the Spring *WEO*.[13]

48. Specifically, we estimate versions of equation (3) where *y* stands for the forecast error of GDP growth as just explained. Estimation was carried out in a panel structure.

$$y_{i,t} = c_{i,t} + u_{i,t} \qquad (3)$$

49. As already noted, the simple average "bias coefficient" for all countries and all years is equal to -0.3, indicating a statistically significant optimistic bias on average over the full sample period. A look at Figure 4, however, gives a more nuanced picture. It shows the average "bias coefficient" year-by-year, including a two standard deviation band. The coefficient is predominantly negative (an optimistic bias) during the 1990s and until 2002. Thereafter it is relatively large and positive for five years before turning negative again in 2008.[14] Looked at in another way, large negative estimates are obtained for years of major regional or global recessions; 1992, 1998, 2001, and 2008 in particular.

Figure 4. Bias coefficients for different years
(In percentage points)



Source: Author's calculations using IMF *WEO*.

[12] The number of countries was dictated by data availability for the entire sample period.

[13] We limit our discussion in this way since our objective here is to illustrate potential pitfalls in interpreting estimation results rather than to provide an exhaustive body of new results.

[14] It should be kept in mind that the values shown in Figure 4 are averages across all 139 countries. Hence, a string of positive or negative forecast errors does not necessarily mean that there is serial correlation in the error of individual countries.

50. Another way to illustrate the importance of the sample period is to explore how the results of Timmermann (2006) hold up when the sample period is extended. Table 3 contains the results of one such exploration. Column 2 shows the averages across country groups of forecast biases measured by the constant in a simple country-by-country regression of the forecast error on a constant for the time period 1990–2003. In line with the results presented by Timmermann, all the averages indicate an optimistic bias.[15]

Table 3. Estimates of bias forecasts of GDP
(In percentage points)

	1990–2003	1990–2011	1990–2011 (without outliers)
Advanced Economies	-0.05	0.07	0.10
Africa	-1.11	-0.77	-0.43
CIS & Mongolia	-1.72	-0.60	0.45
Central & Eastern Europe	-1.08	-0.62	-0.01
Developing Asia	-0.25	-0.02	0.18
Middle East	-0.97	-0.58	0.13
Western Hemisphere	-0.49	-0.24	-0.12

Source: Author's calculations using IMF *WEO*.

51. Column 3 of Table 3 shows the averages for the same countries and country groupings when the sample is extended through 2011. Two changes relative to the calculations for the shorter period are particularly noteworthy: the biases for advanced economies reverse their sign and become pessimistic and the size of all the other biases decreases. In view of what we saw in Figure 3 these changes are not surprising, but they emphasize that conclusions drawn in the literature about the bias of *WEO* forecasts need to be interpreted cautiously, because they are sensitive not only to the time period studied but also to the country groupings considered.

52. Column 4 in Table 3 shows results after the data have been filtered to remove outliers. Hendry and others have argued in a series of papers that removing outliers is appropriate in the context of analyzing model properties, and have developed an algorithm for identifying outliers and breaks in that context.[16] Ericsson (2013) extends these arguments to the evaluation of forecasts. The results for our data once the outliers are removed are striking: the biases become positive (pessimistic) in four out of seven cases. Inspecting the data to uncover when the filter algorithm identified outliers shows that many of these occurred in the regional and global recession periods already signaled in the discussion of Figure 3, namely 1992–94, 1998, 2001, and 2008.

(iv) Efficiency of IMF forecasts

53. If forecast errors are systematically related to information that was available when the forecasts were made, it should be possible to improve the forecasts by making appropriate adjustments. As already noted, Timmermann (2006, 2007) applied this reasoning by asking

[15] The averages for the second column in Table 3 are close, but not identical, to those reported in Table 1 of Timmermann (2006): "Spring current-year forecasts of GDP." We were unable to replicate his results exactly because we could not determine which countries were included in the underlying calculations.

[16] See Hendry and others (2008), Johansen and Nielsen (2009), and Hendry and Santos (2010).

whether *WEO* forecasts took into account information in forecasts of two potential "locomotive" countries, the U.S. and Germany, and put it to the test by regressing forecast errors of GDP growth for other countries on the forecasts of growth in the U.S. and Germany. He argued that if the slope coefficients were statistically significant it could be a sign that international linkages were not taken fully into account in the forecast process.

54. We updated Timmermann's computations using the most recent data. In doing so we added the growth forecasts for China to the regression on the presumption that the Chinese economy has become an increasingly important source of external influence in many countries. The results are in presented in Table 4 in the form of the frequencies with which t-statistics for the slope coefficient were either larger than 2 or smaller than -2. Timmermann recommended that a frequency larger than 0.05 (i.e., 5 percent) be interpreted as a sign of inefficiency in the forecasts. On this basis, our results show that the number of cases of inefficiency has not decreased between the sample used in the original Timmermann study and our extended sample.

55. Following the methodology used by Blanchard and Leigh (2013) it is possible to interpret the coefficient on the forecasts as the difference between the true spill-over from growth in a "locomotive" country and the spill-over implicitly assumed by the IMF forecasters. (See Annex 4.) On this interpretation the IMF has tended to underestimate the spill-overs from the US more frequently than it has tended to overestimate them, since statistically significant coefficients on US forecasts are predominantly positive. The opposite has been the case with respect to China, albeit only slightly so. The IMF has also tended to overestimate the effect of China on the CIS countries and Mongolia, and underestimate its effect in Africa, particularly in the 2000–11 period.

(v) Comparing IMF and consensus forecasts

56. To complete the update of results in the literature using the most recent available data we compare the *WEO* forecast errors for GDP growth with those issued by Consensus Economics. We have already noted that this comparison is the least likely to be influenced by differences in information sets available to the forecasters, provided that the Consensus forecasts are chosen appropriately. We have chosen to compare the Spring *WEO* forecasts with the April Consensus forecasts and the Fall *WEO* forecasts with the September Consensus forecasts. The sample period covers forecasts made in 1990 until 2011 for all countries that are covered by both institutions. We compare the forecasts in terms of their respective root mean square forecast error (RMSFE). In Table 5 we report the ratios of the RMSFE of Consensus forecasts to those of the *WEO*. We also report the results of the Diebold-Mariano test of differences between the two. In each of the four panels, the first (third) column of numbers refers to the number of country/year forecasts in which the RMSFE of Consensus forecasts (*WEO*) is larger. The second and fourth columns in each

panel show the number of cases in which the difference between the accuracy of the two forecasts is statistically significant according to the Diebold-Mariano test.[17]

Table 4. Do *WEO* forecasts sufficiently account for GDP growth from large countries?

	1990–2003						2000–11					
GDP growth rate :	U.S.		Germany		China		U.S.		Germany		China	
T-Statistic :	<-2	>2	<-2	>2	<-2	>2	<-2	>2	<-2	>2	<-2	>2
Spring current-year forecast errors												
Africa	0.04	0.06	0.02	0.10	0.06	0.02	0.02	0.02	0.04	0.00	0.04	0.10
Central/Eastern Europe	0.00	0.40	0.07	0.07	0.00	0.27	0.07	0.60	0.13	0.40	0.00	0.00
CIS and Mongolia	0.00	0.00	0.08	0.31	0.15	0.00	0.00	0.38	0.00	0.23	0.08	0.08
Developing Asia	0.00	0.04	0.09	0.04	0.09	0.00	0.04	0.09	0.17	0.04	0.00	0.09
Middle East	0.06	0.06	0.06	0.00	0.00	0.00	0.19	0.06	0.13	0.06	0.06	0.06
Western Hemisphere	0.00	0.06	0.00	0.00	0.00	0.03	0.00	0.22	0.00	0.06	0.00	0.22
Advanced Economies	0.00	0.28	0.07	0.03	0.00	0.07	0.03	0.14	0.10	0.03	0.00	0.07
Fall current-year forecast errors												
Africa	0.00	0.06	0.04	0.08	0.06	0.02	0.06	0.02	0.06	0.04	0.04	0.08
Central/Eastern Europe	0.00	0.07	0.00	0.00	0.00	0.13	0.07	0.13	0.07	0.07	0.07	0.00
CIS and Mongolia	0.00	0.00	0.00	0.08	0.31	0.00	0.00	0.08	0.00	0.00	0.15	0.00
Developing Asia	0.00	0.04	0.09	0.00	0.00	0.13	0.04	0.04	0.04	0.04	0.04	0.09
Middle East	0.06	0.00	0.06	0.06	0.13	0.13	0.00	0.00	0.06	0.00	0.06	0.13
Western Hemisphere	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.13
Advanced Economies	0.00	0.24	0.03	0.07	0.00	0.03	0.03	0.14	0.07	0.03	0.00	0.00
Spring next-year forecast errors												
Africa	0.02	0.00	0.02	0.12	0.04	0.04	0.02	0.06	0.08	0.02	0.06	0.10
Central/Eastern Europe	0.00	0.00	0.07	0.00	0.00	0.13	0.00	0.20	0.07	0.00	0.13	0.00
CIS and Mongolia	0.00	0.00	0.08	0.54	0.23	0.00	0.00	0.08	0.00	0.00	0.31	0.00
Developing Asia	0.04	0.04	0.09	0.00	0.09	0.04	0.04	0.04	0.13	0.00	0.13	0.04
Middle East	0.00	0.13	0.06	0.00	0.19	0.06	0.00	0.19	0.13	0.13	0.25	0.00
Western Hemisphere	0.09	0.03	0.03	0.03	0.03	0.19	0.00	0.09	0.16	0.00	0.03	0.03
Advanced Economies	0.21	0.03	0.03	0.00	0.00	0.34	0.00	0.07	0.14	0.03	0.17	0.00
Fall next-year forecast errors												
Africa	0.02	0.00	0.02	0.04	0.04	0.04	0.00	0.10	0.04	0.00	0.04	0.10
Central/Eastern Europe	0.00	0.00	0.07	0.00	0.07	0.20	0.00	0.67	0.00	0.13	0.00	0.00
CIS and Mongolia	0.00	0.00	0.00	0.38	0.23	0.00	0.00	0.69	0.00	0.23	0.23	0.00
Developing Asia	0.09	0.04	0.04	0.04	0.04	0.04	0.00	0.17	0.09	0.09	0.04	0.13
Middle East	0.06	0.06	0.00	0.06	0.00	0.00	0.00	0.31	0.00	0.06	0.19	0.06
Western Hemisphere	0.09	0.00	0.03	0.16	0.03	0.03	0.00	0.38	0.00	0.06	0.06	0.03
Advanced Economies	0.07	0.00	0.10	0.00	0.00	0.10	0.00	0.45	0.00	0.00	0.07	0.00

Source: Authors' calculations using IMF *WEO*.

57. Looking first at the totals in each category of forecast, there is little to differentiate between the *WEO* and Consensus in the Spring forecasts, whether they are for current years or the year ahead. For the Fall year-ahead forecasts, the *WEO* appears to have an edge—both in term of the number of cases in which its RMSFEs are lower than those of Consensus forecasts and in terms of the frequency of statistical significance. However, this result is

[17] Country-by-country results can be found in Annex 5.

largely driven by the particular vintage of Consensus forecasts that was chosen for the Fall forecasts.[18]

Table 5. Comparison of forecast errors from Consensus Economics and *WEO*[1]
Ratio of RMSFE (CON/*WEO*)

	Spring Year-Ahead				Fall Year-Ahead				Spring Current-Year				Fall Current-Year			
	WEO (>1)		CON (<1)		*WEO* (>1)		CON (<1)		*WEO* (>1)		CON (<1)		*WEO* (>1)		CON (<1)	
	#	Sig[2]	#	Sig[2]	#	Sig[2]	#	Sig[2]	#	Sig[2]	#	Sig[2]	#	Sig[2]	#	Sig[2]
Africa	1	0	1	0	2	1	0	0	2	1	0	0	2	0	0	0
Asia Pacific	4	0	12	3	8	0	8	0	5	1	11	3	8	2	8	1
Europe	22	4	13	2	29	8	6	0	22	7	13	1	19	1	16	0
Middle East and Central Asia	3	0	3	2	3	1	3	0	3	1	3	1	4	1	2	1
Western Hemisphere	6	1	10	1	10	2	6	0	8	1	8	2	7	3	9	1
TOTAL	36	5	39	8	52	12	23	0	40	11	35	7	40	7	35	3

Source: Authors' calculations using IMF *WEO*.

[1] Sample sizes vary based on available forecasts by country, but typically extend from 1991-2011.Only those countries for which there are at least six observations are included.

[2] Significance indicates that the p-value for the Diebold-Mariano test was less than 0.05.

58. Drilling down by region there are several noteworthy aspects of the results: the *WEO* systematically outperforms Consensus in Africa[19] and in Europe, and the Consensus forecasts typically have lower RMSFEs for the Asia Pacific Region in the case of Spring forecasts.[20]

D. Conclusions

59. Four conclusions emerge from our re-examination of the results in the literature by using the most up-to-date data on forecasts and outturns:

(i) Measures of bias are sensitive to the sample period and country aggregation. This is consistent with survey responses of country authorities reported both in Genberg, Martinez, and Salemi (2014) and in Section II.B above.

(ii) There is something about recessions. The results of bias calculations reported in Figure 4 and Table 3 suggest that the optimistic bias found by a number of authors is the result of particularly large forecast errors that are made in recession periods. Chapter III, Section A below asks how the IMF performs relative to the private sector in this respect.

[18] If the October Consensus forecast vintage is chosen instead of the September vintage then the overall results flip and Consensus appears to have an edge, in terms both of number of lower RMSFEs and of frequency of significance.

[19] Note, however, that there are only two observations for this group of countries.

[20] If the October Consensus forecast vintage is chosen instead of the September vintage then the Consensus forecasts outperform the *WEO* forecasts in Europe for both of the Fall forecasts.

(iii) Comparing forecast accuracy across institutions is fraught with difficulties associated with the timing of the forecasts and therefore with the amount of information that was available when the forecasts were made. When we compare *WEO* forecasts with forecasts from the private sector (specifically Consensus Economics)—where the timing difficulty can be controlled for to some extent—the results do not suggest that either institution systematically outperforms the other.

(iv) The potential efficiency problem highlighted in Timmermann (2006, 2007), whereby information in forecasts for the U.S. and Germany seems not to be fully incorporated in forecasts for other countries, persists in a sample extended to 2011. This could be an indication that interdependencies in the world economy are still not fully taken into account in *WEO* forecasts.

III. EXTENSIONS

58. In this Chapter, we present some new perspectives on IMF forecast accuracy and efficiency. We discuss in turn the influence of severe recessions (or crises) on biases in IMF forecasts; the question whether the experience of the country desk economist matters for the accuracy of forecasts; and how much *WEO* forecasts take into account the interdependencies among economies identified in some of the literature.

A. The Difficulty of Predicting Recessions and Its Relevance for Overall Forecast Evaluation

59. Here we compare the ability of the IMF to forecast recessions with that of the private sector in a sample that covers the years 1991 to 2011. We ask whether the capacity to forecast recessions differs across advanced, emerging, and low-income economies and between program and non-program countries. Finally we assess the characteristics of forecast errors when recession periods are excluded from the sample.

60. As part of their studies of the accuracy of private sector forecasts of output growth, Loungani (2001) and Juhn and Loungani (2002) assessed the ability of private sector forecasters, specifically Consensus Economics, to forecast recessions in a large sample of industrialized and developing countries for the period 1989 to 1999. For the purpose of their assessments, they defined recessions as a negative rate of growth of real GDP.[21] The 2002 paper concluded that "…very few of the 72 recessions that occurred over the sample were predicted a year in advance and two-thirds remained undetected by the April of the year in which the recession occurred" (Juhn and Loungani, 2002, p. 62).

[21] The authors noted that this definition of a recession "encompasses cyclical downturns …, declines in output associated with transition from planned economies to market economies…, and declines associated with crises of various kinds..." (Juhn and Loungani, 2002, p. 58).

61. Applying to our longer sample the same definition of a recession proposed by Juhn and Loungani we confirm their finding that private sector forecasters are not very good at predicting recession. We also find that the Fund's record is no better, although not significantly worse, in this respect than that of the private sector when we restrict our sample to countries for which Consensus forecasts are available. And in an extended sample that includes recession episodes for countries that are not covered by Consensus forecasts we find that the Fund's ability to forecast recessions deteriorates. The decline is most noticeable in regard to low-income countries but is also visible for emerging market economies.

62. Finally we find that when recession episodes are eliminated from the sample, the forecast errors of the IMF do not show the over-optimism that has been found in the literature. This suggests a need for further study of the effect of infrequent but crisis-like episodes of negative economic growth on the Fund's forecast record.

(i) The sample

63. Our sample consists of current-year and one-year-ahead forecasts of real GDP for 1991 to 2011 published in the *WEO* and by Consensus Economics. The *WEO* is typically published in April/May and in September/October, so for the comparison we use Consensus forecasts published in April and September. In view of the production time schedule of the *WEO* (see Genberg, Martinez and Salemi, 2014) this choice is likely to give Consensus Economics a slight advantage in terms of the information that its forecasts can be based on.

64. Following Juhn and Loungani we define a recession as a negative growth rate of real GDP. As the actual value of real GDP in year *t* we use the value published in the Fall issue of the *WEO* in year *t+1*.[22] This process results in a sample of 508 episodes during the period 1991 to 2011 (see Table 6 for a summary and Annex 6 for the full list of episodes).[23] For about one-third of these episodes Consensus forecasts are available allowing for a comparison with IMF forecasts. Note that extending the sample by including recession episodes from the *WEO* database for which there are no Consensus forecasts implies adding mainly emerging economies and low-income countries.

Table 6. Recession episodes, 1991–2011

	Total number of episodes	Advanced Economies	Emerging Economies	Low-Income Countries	IMF Program	Consensus Forecast available
Actual = *WEO* (t+1)	508	82	225	201	97–123	160–174

Source: Authors' calculations using IMF *WEO*.

[22] We also calculated all statistics for an alternative measure of period *t* GDP, namely the value published in the Fall issue of the *WEO* in year *t+2*. None of the results was materially affected by this change. The results are available upon request to the authors.

[23] Forecasts published in the *WEO* are available back to 1990. The first one-year-ahead forecast is therefore for the year 1991.

(ii) Comparing Consensus Economics forecasts with IMF forecasts of recession episodes

65. Results reported in Table 7 show that the findings of Juhn and Loungani still hold in our longer sample: Consensus forecasts predicted very few—8 percent, or 12 out of 156—recessions as of April the year before the recession. This percentage increased to 19 percent as of September the year before, to 71 percent as of April of the recession year, and finally to 92 percent in September of that year. These percentages, which refer to the entire 1991–2010 period, are comparable to, albeit slightly higher than, those reported by Juhn and Loungani for the 1989–99 period.

Table 7. Forecast performance during recession episodes for which Consensus forecasts are available, 1991–2011

		Spring (t-1)	Fall (t-1)	Spring (t)	Fall (t)
Number of countries (observations)			86 (1340–1405)		
Number (share) of recession episodes		160 (12%)	162 (12%)	173 (12%)	174 (12%)
Number (share) of episodes where a recession was forecast (Forecast < 0)	*WEO*	17 (11%)	35 (22%)	119 (69%)	157 (90%)
	Consensus (Mean)	13 (8%)	32 (20%)	123 (71%)	160 (92%)
Number (share) of episodes where forecast was too optimistic (Forecast >Actual)	*WEO*	157 (98%)	159 (98%)	142 (82%)	102 (59%)
	Consensus (Mean)	157 (98%)	158 (98%)	150 (87%)	120 (69%)
Number (share) of recession forecasts that were a false positives (Forecast <0 & No recession)	*WEO*	49 (74%)	45 (56%)	83 (41%)	91 (37%)
	Consensus (Mean)	27 (68%)	41 (56%)	63 (34%)	81 (34%)
Median forecast error (Actual - Forecast)					
Full Sample	*WEO*	-0.02	0.03	0.22	0.21
	Consensus (Mean)	0.10	0.10	0.25	0.22
For all recessions	*WEO*	-5.72	-4.92	-2.04	-0.28
	Consensus (Mean)	-5.70	-5.00	-2.21	-0.65
For non-recessions	*WEO*	0.30	0.31	0.38	0.29
	Consensus (Mean)	0.37	0.35	0.43	0.29

Source: Authors' calculations using IMF *WEO* and Consensus Economics.

66. In comparison, the IMF forecasts performed very similarly. The IMF had a slightly higher success rate in the one-year-ahead forecasts: 10 percent versus 8 percent in the Spring and 21 percent versus 19 percent in the Fall. In view of the potential informational advantage for Consensus Economics, this suggests that the IMF is holding its own relative to the private sector even if on an absolute level the ability to predict recessions is limited. For the current-year projections where the informational advantage is likely to be greater, the record of Consensus forecasts is slightly better than the *WEO* forecasts: 71 percent versus 69 percent for Spring forecasts and 92 percent versus 90 percent for the Fall forecasts.

67. Focusing on instances of false positives, i.e., forecasts of recessions when none occurred, gives similar results as regards the comparison of IMF and Consensus forecasts. For both types of forecasts the incidence of such false positives falls as the horizon of the forecast shortens. The IMF does, however, make a slightly larger number of errors of this kind.

68. Forecasts were too optimistic in most cases, and by a large margin of more than five percentage points for one-year-ahead forecasts as judged by the cross-country median. On this metric, the Fund performed slightly better than the private sector. It is of interest to note that when recession years are taken out, the median forecast errors for the countries in the sample decline substantially for both the IMF and the Consensus forecasts. Furthermore, the forecasts are slightly pessimistic.

(iii) Results for all countries in the *WEO*

69. Table 8 presents the results for all the countries included in the *WEO* database. The following five conclusions can be drawn from this table. First, for year-ahead forecasts, the share of recessions predicted is still small, and similar to that for the restricted sample in Table 7. For current-year forecasts it is substantially smaller, and in fact surprisingly small for the Fall forecasts. Second, in terms of the share of the forecasts that were too optimistic, the results for the full and the restricted sample are similar. Third, the incidence of false positives is substantially smaller for current-year forecasts and somewhat smaller for year-ahead forecasts. Fourth, the size of forecast errors is slightly larger for year-ahead forecasts but substantially larger for current-year forecasts. Fifth, when non-recession years are excluded there is little evidence of over-optimism in IMF forecasts; indeed, as measured by the median forecast errors, all IMF forecast errors are either very close to zero or slightly positive, pointing towards a slight pessimism in the forecasts.

Table 8. Forecast performance during all recession episodes, 1991–2011

	Spring (t-1)	Fall (t-1)	Spring (t)	Fall (t)
Number of countries (observations)	185 (3885)			
Number (share) of recession episodes	508 (13%)			
Number (share) of episodes where a recession was forecast (Forecast < 0)	59 (12%)	88 (17%)	254 (50%)	345 (68%)
Number (share) of episodes where forecast was too optimistic during recession (Forecast > Actual)	459 (90%)	455 (90%)	416 (82%)	336 (66%)
Number (share) of recession forecasts that were a false positives (Forecast <0 & No recession)	108 (65%)	111 (56%)	171 (40%)	198 (36%)
Median forecast error (Actual - Forecast)				
Full sample	-0.29	-0.20	0.00	0.00
Recessions Only	-6.27	-5.73	-3.72	-1.49
Excluding Recessions	0.00	0.02	0.10	0.08

Source: Authors' calculations using IMF *WEO*.

(iv) Does a country's stage of economic development matter?

70. The full *WEO* sample contains a large number of emerging economies and low-income countries that have experienced recession episodes. As the results in Table 9 show, separating these countries from the advanced economies yields additional insights into the difficulties of forecasting recessions. For advanced countries the record is very similar to that reported in Table 7 where comparisons with Consensus forecasts were made. Recessions are

not forecast one year ahead for the most part, but once the economy is already in recession, the recognition that this is the case improves substantially.

Table 9. Forecast performance during recession episodes, 1991–2011, by country development level

	Spring (t-1)	Fall (t-1)	Spring (t)	Fall (t)
Advanced Economies				
Number of countries (observations)		34 (606)		
Number (share) of recession episodes		82 (14%)		
Number (share) of episodes where a recession was forecast (Forecast < 0)	13 (16%)	27 (33%)	65 (79%)	79 (96%)
Median forecast error (Actual - Forecast)				
Full sample	-0.30	-0.19	0.07	0.10
For all recessions	-4.14	-3.64	-0.90	-0.16
For non-recessions	0.00	0.12	0.15	0.11
Emerging and Developing Economies				
Number of countries (observations)		94 (1711)		
Number of recession episodes		255 (13%)		
Number (share) of episodes where a recession was forecast (Forecast < 0)	28 (12%)	34 (15%)	122 (54%)	161 (72%)
Median forecast error (Actual - Forecast)				
Full sample	-0.04	0.00	0.06	0.10
For all recessions	-6.50	-5.92	-3.38	-1.44
For non-recessions	0.30	0.30	0.30	0.23
Low-Income Countries				
Number of countries (observations)		80 (1568)		
Number of recession episodes		201 (13%)		
Number of episodes where a recession was forecast (Forecast < 0)	18 (9%)	27 (13%)	67 (33%)	105 (52%)
Median forecast error (Actual - Forecast)				
Full sample	-0.50	-0.42	-0.20	0.00
Recessions Only	-7.03	-6.89	-5.00	-3.59
Excluding Recessions	-0.11	-0.08	0.00	0.00

Source: Authors' calculations using IMF *WEO*.

71. For emerging economies and especially for low-income countries the situation is different. While year-ahead forecasts continue to miss recessions in a large percentage of cases, the forecast performance does not improve substantially in current-year forecasts. Indeed, only about half of the recession episodes are recognized as such in the Fall of the year of the recession. There are at least three plausible explanations for this. First, it is possible that a number of recessions in low-income countries are due to factors that cannot be forecast even during the year that they occur. Second, in a low-income country data are likely to arrive with longer lags, in addition to being less reliable. Consider a situation where there is a one-year delay in the arrival of reliable information about current economic conditions. In this case a forecaster in period *t* is effectively constrained to work with information from the year before. In effect, the current-year forecast in such an economy is like a year-ahead forecast in an economy where information is up to date. In this perspective the forecast

performance for low-income countries may reflect less the ability of the forecaster than the lack of information. From the point of view of being able to foresee economic troubles and propose appropriate policy responses, there seems to be a high payoff to improving data availability in these cases. It is possible that the relatively poor record in forecasting recessions in low-income countries could be the consequence of a lack of attention and resources allocated to developing forecast and early-warning methodologies specific to these types of economies.

(v) Are program cases special?

72. Table 10 contains results pertaining only to IMF program cases. Two features of these results are worth highlighting. First, the frequency of recession episodes that are forecast in these cases is similar to that observed for emerging markets and low-income countries in general—a finding that may be related to the fact that many program cases are drawn from these two categories. Second, once outright recessions are excluded from the calculations there is no sign of a significant bias in forecasts for program cases.

Table 10. Forecast performance during recession episodes, 1991–2011: IMF program countries

	Spring (t-1)	Fall (t-1)	Spring (t)	Fall (t)
Number of countries (observations)	115 (919-988)			
Number (share) of recession episodes	97 (11%)	123 (12%)	97 (11%)	123 (12%)
Number (share) of episodes where a recession was forecast (Forecast < 0)	14 (14%)	27 (22%)	41 (42%)	77 (63%)
Number (share) of episodes where forecast was too optimistic during recession (Forecast > Actual)	96 (99%)	114 (93%)	88 (91%)	87 (71%)
Number (share) of recession forecasts that were a false positives (Forecast <0 & No recession)	16 (53%)	23 (46%)	22 (35%)	33 (30%)
Median forecast error (Actual - Forecast)				
Full sample	-0.43	-0.30	-0.05	0.00
Recessions Only	-7.03	-6.93	-4.94	-1.94
Excluding Recessions	-0.07	0.00	0.00	0.06

Source: Authors' calculations using IMF *WEO*.

(vi) Are IMF forecasts of real GDP growth over-optimistic?

73. Evidence reviewed elsewhere in this paper and in Freedman (2014) frequently suggests that IMF growth forecasts tend to be too optimistic, by predicting output growth higher than the actual outturn. The results in Tables 9 and 10 provide a perspective on these results that has not been emphasized in the literature: for virtually all categories of countries and forecasts they show no sign of an optimistic bias when recession years are excluded from the calculations. This finding is not necessarily just a mechanical implication of excluding observations with large over-predictions of growth: if an optimistic bias were a regular feature of IMF forecasts, it should be evident also in non-recession years.

74. Viewed in this light, the over-optimism often ascribed to IMF forecasts appears to be due largely to an inability to foresee crises and negative growth. This inability is shared by private sector forecasters. But in view of the Fund's role to monitor the stability of the international monetary system, perhaps one should be entitled to expect more from the IMF in this respect.

(vii) Why are recessions not forecast?

75. While it is clear that some events may be unpredictable, Juhn and Loungani (2002) argue that forecasters' inability to predict recessions could be "either because forecasters lack the requisite information (in terms of reliable real-time data or models) or because they lack the incentives to predict recessions" (p. 63).

76. There is indeed evidence suggesting that IMF staff may lack incentives to predict recessions, because the internal forecasting process at the IMF does not encourage forecasts that "rock the boat." As part of the review process, as explained in Section III.B below, staff forecasts are checked against those of other forecasters and need to be justified if they are different. Thus, a desk economist can minimize the amount of scrutiny his/her forecasts will get by not differing significantly from the consensus. And while this scrutiny operates symmetrically, the cost of forecasting a recession that does not materialize may be perceived as being higher than the cost of having wrongly predicted a boom.

77. There is also substantial evidence supporting the idea that a lack of information hinders the ability to forecast recessions. As seen above, the higher success rate of forecasting recessions in advanced economies, compared to forecasting recessions in low-income countries, is likely related to differences in the availability of high frequency data and the use of more sophisticated and complex models. In fact, recent research done both inside and outside the IMF shows that the use of leading indicators can improve IMF forecasts both in and out of recessions.[24] Some area departments in the IMF are already working to incorporate more high frequency data from more unconventional sources into their forecasts.

B. IMF Forecast Accuracy: Does Staff Experience Matter?

78. The IMF's *WEO* forecasts are constructed using a bottom-up approach. During the *WEO* forecasting process every IMF country desk produces a set of forecasts for its country. Because the IMF membership is so diverse, the staff of each country desk are given substantial autonomy over how they produce their forecasts and in practice the forecast methods they use vary substantially. From interviews and a survey of IMF country desk staff, it is clear that all country desk forecasts rely at least a partly on the judgment of the desk economist (see Genberg, Martinez and Salemi, 2014).

[24] Baba and Kisinbay (2011) show how the use of leading indicators can help identify recessions, while Drechsel and others (2013) show that models that include leading indicators can outperform IMF forecasts at various horizons.

79. Another important feature of the IMF's *WEO* forecasting process is that it employs a top-down coordination process to help ensure global and regional consistency. The forecasts made by desk economists are reviewed and agreed upon by the country team and have to be approved by the mission chief as well as the area department. They are also reviewed by the Research Department, and the country authorities are given opportunities to comment prior to publication. These checks may help correct for inconsistencies among individual country desk forecasts and mitigate the impact of any misjudgment on the part of the country desks.

80. One way to ascertain how well the overall *WEO* forecasting process works is to examine whether individual staff judgment discernibly affects the accuracy of forecasts. Since the use of judgment in forecasts is not always directly measurable, it is necessary to gauge its effect on forecast accuracy in some other way. Judgment is influenced by desk economists' specific knowledge of and experiences in the country for which they produce their forecasts. Staff judgment is also influenced by past training and experiences in other countries and at the IMF in general.[25] By examining the stock of country-specific and general experience that desk economists have when they produce their forecasts, it may be possible to understand how informed the judgment in their forecasts is.

81. Using a combination of findings from statistical analysis, survey results, and interviews, this section examines how staff experience affects forecast accuracy. We use a unique internal IMF dataset to compare *WEO* short-term GDP forecast errors for a large set of countries over the past five years against the experience levels of the desk economists who produced those forecasts. In doing so, we attempt to analyze whether and how different types and levels of experience affect the forecast errors, and if so, whether this effect changes across different groups of countries and different forecast horizons. We reinforce our analytical findings with results from our staff survey and from interviews with IMF staff. Subsection (i) provides a review of the relevant literature. Subsection (ii) describes the methodology and the dataset and subsection (iii) presents the statistical analysis. Subsection (iv) provides some additional survey results as well as some findings from interviews with IMF staff. Subsection (v) concludes and presents some recommendations.

(i) Forecasters' experience and performance

82. Numerous studies, focusing mainly on security analysts' forecasts of corporate quarterly earnings, find a relationship between forecaster experience and forecast accuracy. The earliest of these, Mikhail and others (1997) controls for what it calls "innate" and "environmental" sources of forecast performance, and finds a "statistically significant decline in the absolute value of analyst forecast errors as firm-specific experience increases"

[25] Judgment can also be influenced by experience and training outside the IMF. Due to data limitations we limit our examination to experiences within the Fund and therefore implicitly assume that on average desk economists join the IMF with the same level of experience and with similar training.

(p. 132).[26] Clement (1999) extends these findings to both firm-specific and general experience, and Clement and others (2003) and Bollinger (2004) extend the analysis to additional countries and find a significant (negative) relationship between firm-specific experience and forecast errors.[27] However, after controlling for analyst-specific effects, Jacob and others (1999) find that improvements in forecast performance are mainly attributable to the fact that "more capable analysts survive longer and are thereby over-represented among the observations with high values of experience" (p. 71).

83. It is an open question whether the literature regarding non-environmental influences on forecast accuracy is applicable to the IMF. Private firms where forecast performance is often directly related to job performance contrast with the IMF, where forecasts serve as an input to policy advice and their accuracy is not considered in a desk economist's annual performance review. Similarly, while private analyst forecasts often produce their forecasts independently, IMF desk economists' forecasts (in particular those published in the *WEO*) go through several rounds of reviews and cross-checks, as noted above (Genberg, Martinez and Salemi, 2014).

84. Studies concerning IMF forecast accuracy have largely focused on "environmental" sources. There is a substantial literature on the effect on the forecasts when a country has an IMF program (see for example Musso and Phillips, 2002; Baqir and others, 2005; and Atoyan and Conway, 2011). Others have argued that the quality of the data provided by country authorities can affect forecast performance (Tong, 2004 and Mrkaic, 2010). These studies make clear that environmental factors can have a considerable effect on the accuracy of the IMF's forecasts, though their findings do not exclude the possibility that non-environmental effects also play a role.

85. Other evidence suggests that desk experience could affect IMF forecast performance. The IEO has repeatedly found that high turnover hinders the IMF staff's ability to build country-specific expertise as well as their performance in a variety of tasks (IEO, 2003; 2009; 2011; 2013). While the IEO studies have not focused on forecasts specifically, these repeated findings do beg the question whether staff's country experience affects their forecast performance.

86. Further evidence from interviews with senior staff portrays a complicated relationship between desk experience and forecast performance. One senior staff member mentioned that less experienced desk economists often make simple mistakes when producing their submissions to the *WEO.* In contrast, other senior staff thought that newly assigned staff might be more likely to review and improve the forecast methodology, which may have

[26] Innate factors are defined as industry forecasting experience and general forecasting experience. Environmental factors are defined as the number of analysts following a company and the amount of information available to the analysts.

[27] Garcia-Meca and Sanchez-Ballesta (2006) appear to confirm these results as well but find that the relationship between experience and forecast errors is strongest in the U.S.

grown stale under previous incumbents. These views provide somewhat different perspectives on whether experience on a given country desk may or may not improve forecast accuracy and illustrate the importance of clarifying whether there is a relationship between staff experience and forecast accuracy at the IMF.

(ii) Data and methodology

87. Our computation of the *WEO* forecast errors uses the actual value for time *t*, assumed here to be the value for *t* at time *t+2*, and the forecast for time *t*. The dataset contains forecasts from 1991 through 2011 for 184 countries in the *WEO* database (see Annex 7 for the list of countries). The forecast for a given country is subtracted from the actual value to obtain the absolute forecast error:

$$WEO_{ist} = | A_{ist} - F_{ist} |, \quad (4)$$

where WEO_{ist} denotes the absolute *WEO* forecast error for country *i* during season *s* and year *t*, *Aist* denotes the "actual" real GDP growth rate, and F_{ist} denotes the forecast of the real GDP growth rate. The absolute error is used since this analysis focuses on the magnitude of the error and on not the direction in which it occurred. While the squared forecast error is a more common focus in the literature, the absolute forecast error has several benefits in that it allows for a clear interpretation of the regression coefficients and does not emphasize large outliers.

88. The main explanatory variable used to capture country-specific experience comes from a dataset that tracks staff country responsibilities. The database was established in 2002 by the IMF Human Resources Department (HRD) and relies on reporting from the area departments. Because the area departments reported only intermittently before 2005, and because inconsistencies arose during the IMF restructuring in 2007, our analysis here only uses the database from 2007 through 2011.[28] It is important to note that by restricting analysis to this period, the results may suffer from the fact that much of the Fund's senior staff experience was lost following the IMF downsizing in 2007 and that the global financial and economic crisis prevailed throughout the entire period. Given these caveats, it is unclear how far the results from this analysis can be generalized.

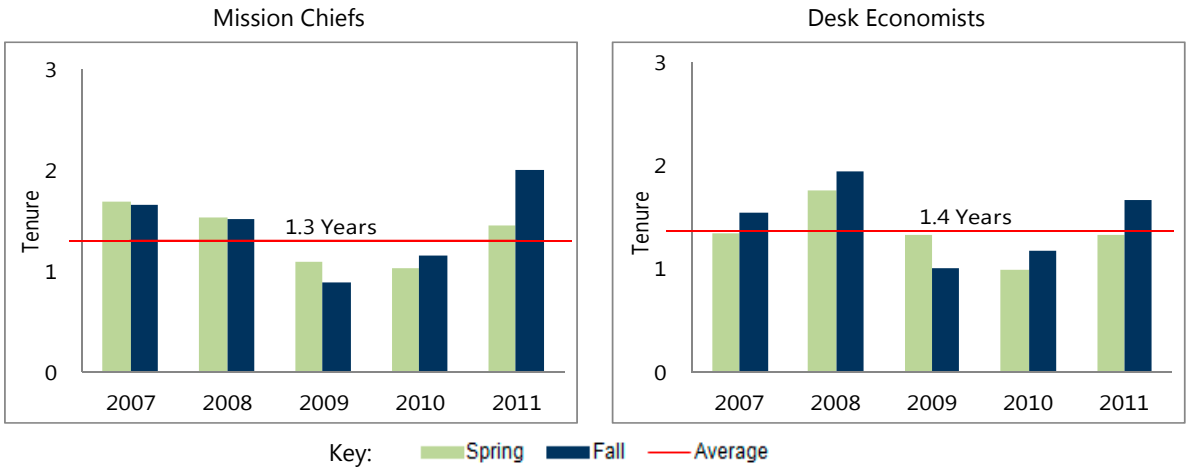
89. Additional data issues remain. For example, occasionally individuals enter the database and are not removed when they leave the country desks or the IMF.[29] To address these issues, the HRD dataset is combined with another dataset on IMF mission travel from the Office of Budget & Planning (OBP). As a result, only those individuals who appear in

[28] Parts of this analysis use data from prior to 2007, such as how many country desks an individual has worked on.

[29] Moreover, some individuals are recorded as not having worked on a country at all. There is little that can be done to correct this problem and therefore it is possible that there is a bias in the results.

both the OBP and HRD databases are included in the data.[30] The final dataset contains tenure information on 541 IMF staff members who occupied 1,466 desk economist positions and 779 mission chief positions. The average tenure of the desk economists at a country desk in the data set between 2007 and 2011 is 1.4 years while the average tenure for mission chiefs is 1.3 years (see Figure 5).[31]

Figure 5. Average length of staff tenure on IMF country desks when *WEO* forecasts are produced[1]
(In years)



Source: Authors' calculations.
[1] Note: Sample size varies by role, year, and season.

90. The next step is to calculate how long each mission chief and desk economist has worked on a given country when *WEO* forecasts are made. Since all that is known is when the *WEO* forecasts are released, rather than when they are finalized, it is assumed that the forecasts are produced approximately one month prior to the release date. Thus, the Fall *WEO* forecast production date is assumed to be August 30 while the Spring *WEO* forecast production date is assumed to be March 30.[32]

91. Given the *WEO* forecast production date, one can calculate how long a desk economist or mission chief has been working on the desk prior to the *WEO* forecast. The

[30] When there is a discrepancy of more than a year between when an individual started/completed their last mission on a given country and when their desk tenure started/ended, the analysis relies on the OBP data. Several manual checks were performed and this appears to correct for most of the issues in the HRD dataset. It is unclear however, whether the potential exclusion of desk economists who do not appear in the travel dataset biases the results or not. This is another reason for starting the analysis after 2007 given that the IMF restructuring exercise reduced the number of economists participating in IMF staff visits.

[31] This corresponds roughly with IMF (2013), which found an average tenure of 1.3 years for both mission chiefs and desk economists as of April 2012.

[32] Based on past *WEO* schedules, the first deadline for all country forecasts can be up to two months before the release date. However, country desks often revise their forecasts up to two weeks prior to the release date.

average tenure of all the economists working on the desk when the forecast was produced is calculated since multiple economists may work on a given country desk, and there is no information available on which economist was responsible for the GDP forecasts.[33] Thus, the calculation of the average desk economist tenure at the time of the forecast is as follows:

$$DE_{ist} = \frac{\sum_{j=1}^{DNist} WEOF_{st} - DESKS_{jist}}{DN_{ist} * 365.242}, \quad (5)$$

where $WEOF_{st}$ is the *WEO* forecast date for season *s* and year *t*, $DESKS_{jist}$ is the start date for each individual economist *j* working on a country *i* during season *s* and year *t*, DN_{ist} is the number of economists working on a country *i* for season *s* and year *t*, and DE_{ist} is the average desk economist tenure in years for country *i* during season *s* and year *t*. A similar calculation is also made for mission chiefs.

92. In addition to country-specific experience, the analysis also takes into account general forecasting and/or IMF experience. The variables that are included to capture this are (all averaged over the number of desk economists working on a country at a given time): tenure at the IMF, the number of previous country desk assignments, the total number of desk economists working on the country desk at a given time, and the number of weeks of IMF forecast-related training attended.[34] Each of these variables tries to capture a different aspect of a desk economist's general work and forecasting experience that could influence the accuracy of their forecasts.

93. The analysis also attempts to control for other "environmental" influences on forecast performance. Based on findings in the literature, dummy variables are included for countries that subscribe to the IMF's data standards,[35] or have an IMF program, and for level of development, geographical region, the IMF downsizing in 2007, and for the 2008/2009 global financial crisis. By controlling for these different aspects the analysis ensures that the results are not influenced by potential missing variables.

(iii) Analysis and results

94. The initial analysis uses a panel least-squares model with heteroskedasticity-corrected standard errors (White); approximately 149 country cross-sections (depending on the

[33] By taking the average it is assumed that the experience of the newer desk economists will balance out the experience of the older ones. Alternatively one could assume that there is a perfect flow of information between the economists which means that one should focus on the desk economist with the most experience. Or one could assume that the forecasts are only as good as the weakest link, which would imply looking at the newest economist on the country desk. These alternative assumptions do not appear to have a significant impact on the results.

[34] For more information on the types of IMF forecast-related training courses included in this analysis see Luna (2014a).

[35] Special Data Dissemination Standard (SDDS) or General Data Dissemination Standard (GDDS).

sample); and annual data from 2007 to 2011. Random cross-section country effects are applied and time dummy variables are included.[36] The initial baseline equation is:

$$WEO_{ist} = \beta_0 + \beta_1 * ME_{ist} + \beta_2 * DE_{ist} + \beta_3 * GE_{ist} + \beta_4 * X_{ist} \quad (6)$$

where WEO_{ist} is the same as in (4), ME_{ist} and DE_{ist} are derived from (5), GE_{ist} are the additional variables concerning general IMF and forecasting experience, and X_{ist} are the additional control variables. Table 11 presents the results of equation (6) for the various *WEO* forecast horizons and seasonal vintages. The constant term, the dummy variable for advanced economies, and the time dummy for the year 2009 are significant across all horizons and vintages.

95. Not surprisingly, as the forecast horizon becomes shorter the model is less able to explain the *WEO* forecast errors. Since the Fall current-year forecast is made a few months before the end of the year, it becomes less and less likely that environmental factors and/or judgment will affect forecast accuracy. On the other hand, the Spring next-year forecasts are produced almost two full years prior to the date they are trying to forecast—which means that environmental factors and/or judgment could potentially play a significant role, as is evidenced by the fact that significance is much higher across all variables.

96. In general, the regression results support the hypothesis that both country-specific and general forecasting experience improves forecast accuracy. While desk economist tenure is only significant in the Fall year-ahead forecast, the coefficient is almost always negative across the forecast vintages, suggesting that each additional year a desk economist spends on a country desk reduces the forecast errors he/she will make. In terms of general experience, there are consistently negative relationships between the forecast errors and how long a desk economist has been at the IMF, the number of previous desk assignments, and the amount of IMF forecast-related training that a desk economist has attended.[37] However, not all types of staff experience are associated with a reduction in forecast errors: the results for mission chief tenure are ambiguous in that they are never significant and switch signs across the different forecast vintage samples.

[36] The Hausman test was used in order to determine whether to use fixed or random country effects.

[37] This does not necessarily imply that participation in more IMF forecast-related training results in smaller forecast errors. It is possible that desk economists who are already good at forecasting may attend more IMF forecast-related training for their own interest or as a way to maintain their skills.

Table 11. Determinants of *WEO* forecast errors[1]

Dependent Variable:	Absolute *WEO* Forecast error of real GDP growth			
Time Period:	2007–11			
Horizon:	Year Ahead		Current Year	
Vintage:	Spring	Fall	Spring	Fall
Constant	4.043***	3.992***	1.974**	1.502**
	(1.052)	(0.820)	(0.848)	(0.602)
Mission Chief Tenure	-0.0250	0.118	0.0480	-0.0654
	(0.161)	(0.114)	(0.119)	(0.0760)
Desk Economist Tenure	-0.177	-0.544***	0.0127	-0.0839
	(0.154)	(0.122)	(0.0770)	(0.0824)
IMF Tenure	-0.0553	-0.00980	-0.0276	-0.0178
	(0.0435)	(0.0349)	(0.0281)	(0.0162)
# of Previous Desk Assignments	-0.139	-0.253*	-0.165	-0.0520
	(0.142)	(0.130)	(0.126)	(0.0689)
# of Country Desk Economists	-0.349**	0.0510	-0.219*	0.0196
	(0.160)	(0.146)	(0.116)	(0.0692)
IMF Forecasting-Related Training	-0.471**	-0.483***	-0.114	-0.0533
	(0.214)	(0.165)	(0.132)	(0.0968)
SDDS Subscribers	0.542	0.0887	0.235	-0.307
	(0.515)	(0.448)	(0.517)	(0.381)
GDDS Subscribers	-0.183	-0.362	-0.0305	-0.00968
	(0.476)	(0.474)	(0.483)	(0.372)
IMF Program	-0.440	-0.102	0.382*	-0.0175
	(0.334)	(0.319)	(0.224)	(0.154)
Advanced Economy	-1.472***	-1.124***	-0.802**	-0.392**
	(0.519)	(0.411)	(0.356)	(0.198)
Low-Income Country	-0.367	-0.0158	0.165	0.0486
	(0.393)	(0.364)	(0.363)	(0.273)
Africa	-0.505	-0.546	-0.502	-0.221
	(0.564)	(0.511)	(0.555)	(0.356)
Asia Pacific	-0.189	-0.670	0.551	0.149
	(0.630)	(0.511)	(0.523)	(0.257)
Middle East and Central Asia	-0.162	-0.296	-0.0363	-0.231
	(0.702)	(0.485)	(0.495)	(0.282)
Western Hemisphere	-0.321	-0.170	-0.142	-0.0194
	(0.491)	(0.379)	(0.332)	(0.200)
Year 2007	-0.160	-0.752*	-0.0104	0.0880
	(0.397)	(0.403)	(0.230)	(0.251)
Year 2008	0.263	-0.0578	0.672**	0.254
	(0.382)	(0.403)	(0.300)	(0.264)
Year 2009	3.419***	2.760***	1.126***	0.154
	(0.564)	(0.484)	(0.305)	(0.227)
Year 2010	1.374***	0.272	0.477*	0.0918
	(0.419)	(0.381)	(0.274)	(0.238)
R^2	0.274	0.250	0.122	0.0931
Sigma	2.403	2.041	1.391	0.828
Number of Observations	351	333	291	256
Number of Countries	141	138	131	129

Source: Authors' calculations.

[1]Robust standard errors reported in parentheses. *** denotes significance level at 1 percent; **, at 5 percent, and *, at 10 percent.

97. The results on IMF programs and forecast performance are mixed. Only the Spring current-year forecast vintage provides an indication that having an IMF program is positively associated with larger forecast errors. This is somewhat surprising given that previous literature has in general found that forecasts for countries with IMF programs generally suffer from an optimistic bias. It is important to note that our analysis looks only at the absolute size of the errors, whereas previous studies focused on directional biases. It is still possible for the forecasts for countries that have IMF programs to be biased in a specific direction without the errors being larger than for countries without an IMF program. The use of year-fixed effects in the analysis may also capture some of the program effects.

98. The results do not seem to suggest that a country's IMF data subscription causes an improvement in *WEO* forecasting performance. Contrary to Tong (2004) and Mrkaic (2010), we do not find a significant relationship between the *WEO* forecast errors and whether a country subscribes to the IMF's SDDS or GDDS. These results may be influenced by the relatively short time period we cover (between 2007 and 2011 there were relatively few changes in the membership of these systems) and by other structural changes which may overwhelm the results. Similarly, the dummy variables for whether a country subscribes to SDDS or GDDS may not sufficiently capture the differences in statistical quality and availability across countries.[38]

99. The next step is to remove variables that were unable to explain the *WEO* forecast performance across the different forecast vintages to see whether the results are robust to these changes. The following variables are dropped from the analysis: mission chief tenure, SDDS and GDDS subscription, and IMF regional department. Additionally, when the relationship is plotted between desk experience and forecast errors across different levels of economic development (see Figure 6), the slopes suggest there are differences across country types. As a result, the analysis is extended to allow for interactions between desk economist tenure and different levels of economic development. These changes also allow for the sample to be extended to a larger set of countries.[39]

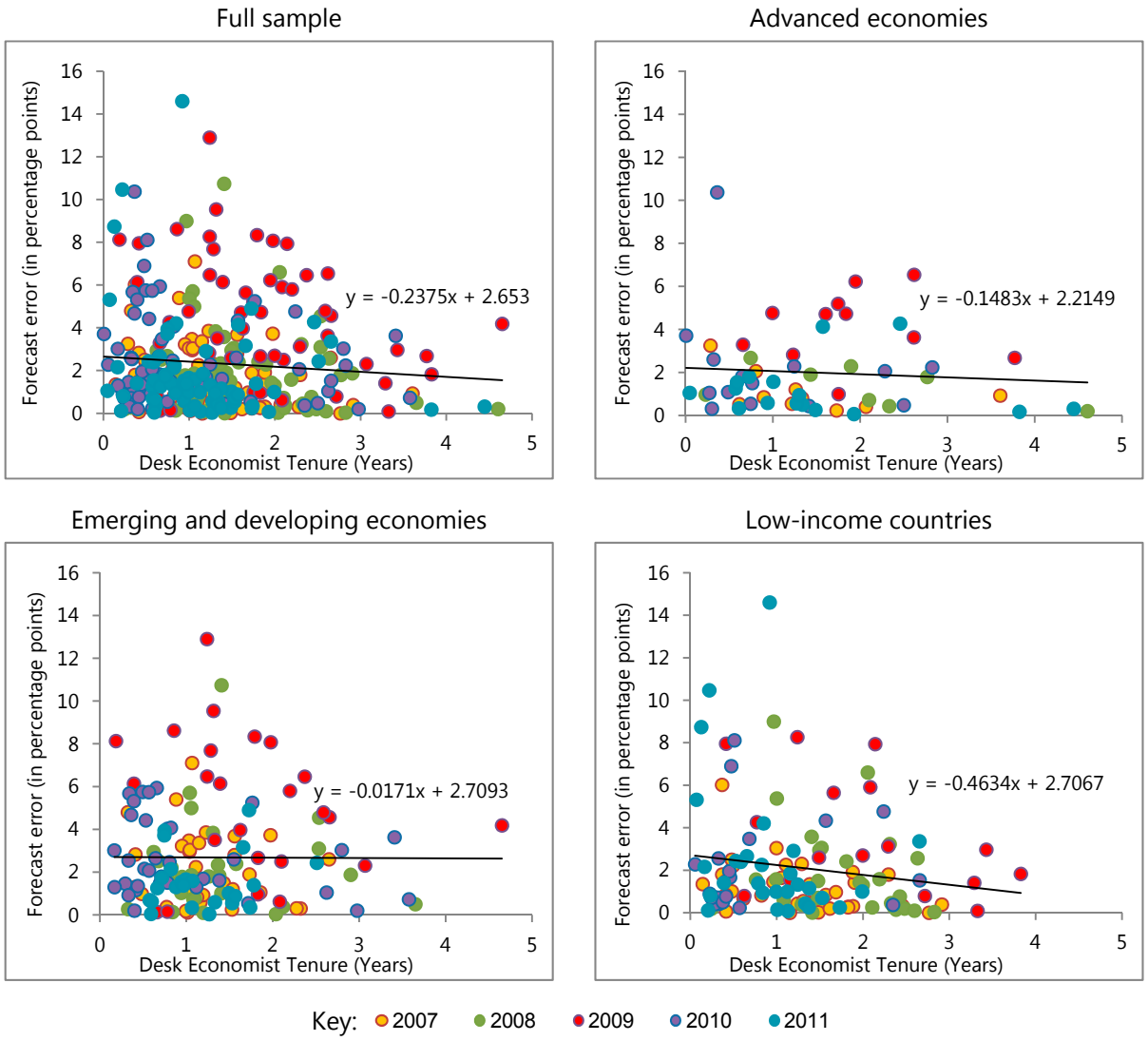
100. Table 12 presents the results of the revised analysis. The results suggest that the relationship between desk economist tenure and *WEO* forecasts is particularly strong and significant for low-income countries. In the year-ahead forecasts, the evidence suggests that each additional year of country-specific experience is associated with a decline in the forecast error of real GDP growth by more than half a percentage point. Interestingly, for advanced economies there is often a negative, albeit minuscule, relationship between the forecast errors and country-specific experience. For emerging and developing economies the

[38] It is possible that the dummy variables for country development level are better than the IMF data subscription dummy variables at picking up these differences, especially given that there is a high correlation between the advanced-economy dummy and the SDDS-subscription dummy.

[39] The sample of countries therefore expands from 149 to 167 countries.

relationship between forecast errors and desk experience tends to be positive, although never significant.

Figure 6. *WEO* GDP forecast errors and desk economist tenure[1]



Source: Authors' calculations.

[1] Figure uses the Fall Next Year GDP forecast errors.

Table 12. Determinants of *WEO* forecast errors, extended sample[1]

Dependent Variable:	Absolute *WEO* Forecast error of real GDP growth			
Time Period:	2007–2011			
Horizon:	Year Ahead Forecast		Current Year Forecast	
Vintage:	Spring	Fall	Spring	Fall
Constant	3.090***	2.922***	2.816***	1.474***
	(0.574)	(0.574)	(0.623)	(0.448)
Advanced * Desk Economist Tenure	-0.0123	-0.0843	-0.0605	-0.0327
	(0.179)	(0.207)	(0.0956)	(0.0464)
Emerging * Desk Economist Tenure	0.0718	0.0376	0.00270	0.0114
	(0.164)	(0.174)	(0.137)	(0.0713)
Low Income * Desk Economist Tenure	-0.548***	-0.706***	-0.124	-0.122
	(0.206)	(0.228)	(0.107)	(0.114)
IMF Tenure	-0.0149	0.0421	-0.0117	-0.0166
	(0.0288)	(0.0400)	(0.0277)	(0.0137)
Number of Previous Desk Assignments	-0.143	-0.271**	-0.181*	-0.0562
	(0.112)	(0.134)	(0.105)	(0.0646)
Number of Country Desk Economists	-0.246*	-0.0764	-0.266**	0.0414
	(0.134)	(0.117)	(0.118)	(0.0580)
IMF Forecasting-Related Training	-0.285*	-0.384***	-0.313***	-0.139*
	(0.157)	(0.147)	(0.119)	(0.0709)
IMF Program	-0.283	-0.376	0.190	-0.0217
	(0.227)	(0.295)	(0.253)	(0.166)
Advanced Economies	-0.823**	-0.932**	-0.758**	-0.471***
	(0.367)	(0.455)	(0.361)	(0.181)
Low Income Countries	-0.0192	0.629	-0.168	0.518**
	(0.422)	(0.444)	(0.350)	(0.235)
Year 2007	0.0866	-0.312	-0.144	-0.399
	(0.330)	(0.311)	(0.326)	(0.324)
Year 2008	0.460	0.225	0.283	0.126
	(0.303)	(0.330)	(0.339)	(0.342)
Year 2009	3.699***	3.095***	0.566*	-0.185
	(0.435)	(0.421)	(0.300)	(0.332)
Year 2010	1.160***	0.374	0.155	-0.340
	(0.326)	(0.303)	(0.308)	(0.342)
R^2	0.243	0.218	0.0610	0.106
Sigma	2.383	2.186	1.607	1.014
Number of Observations	570	546	490	439
Number of Countries	166	163	160	151

Source: Authors' calculations.

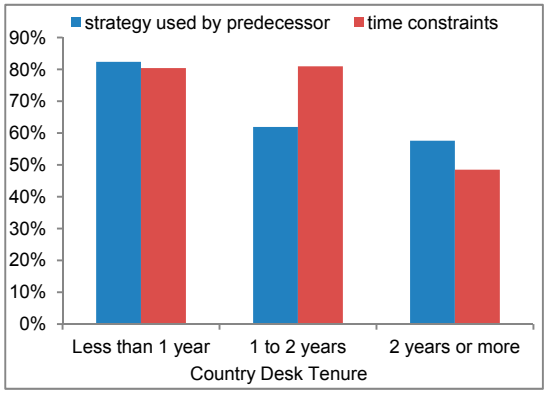
[1]Robust standard errors reported in parentheses. *** denotes significance level at 1 percent, ** at 5 percent, and * at 10 percent.

(iv) Additional evidence from surveys and interviews

101. In this section, the findings of the statistical analysis presented above are cross-checked and supplemented through the results of the IEO survey of country desk economists and through follow-up interviews with a random sample of IMF staff. In general the results remain largely consistent across the three sources of information and provide a relatively clear picture of how desk economists at the IMF incorporate experience and judgment into their forecasts.

102. The survey results of IMF country desk economists suggest that country-specific experience plays a role in how forecasts are produced at the IMF. When economists join a country desk they rely on the forecasting methods and approaches used by their predecessors (Figure 7). This is largely because desk economists have to "learn on the job" when they first start at a country desk. However, as a desk economist becomes more familiar with the structure of the economy, he/she starts to take on more ownership of how the forecasts are produced and to rely less on the predecessor's work. As one desk economist said, "[a]t the beginning [it is] very useful to rely on what is there, while you learn [about] the economy, only then can you think of improving [the forecasts]." Thus, over time the changes that desk economists make to the process often lead to improvements in forecast accuracy.

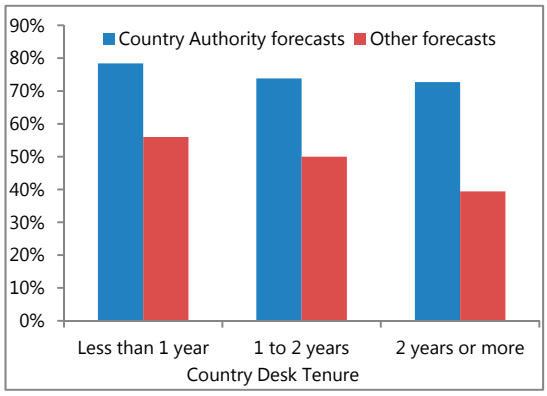
Figure 7. How important were the following in your choice of forecast methods?
(Percent who answered Very / Somewhat important)



Source: IEO Forecast Evaluation Survey.

103. As well as becoming less dependent on their predecessors' forecast methodologies, desk economists who work on the same country desk for a longer period also tend to rely less on forecasts from external sources (Figure 8). This suggests that more experienced desk economists have more confidence in their understanding of the economy and in their ability to produce an accurate forecast for that country.

Figure 8. How important were the following to the production of your country forecasts?
(Percent who answered Very / Somewhat important)



Source: IEO Forecast Evaluation Survey.

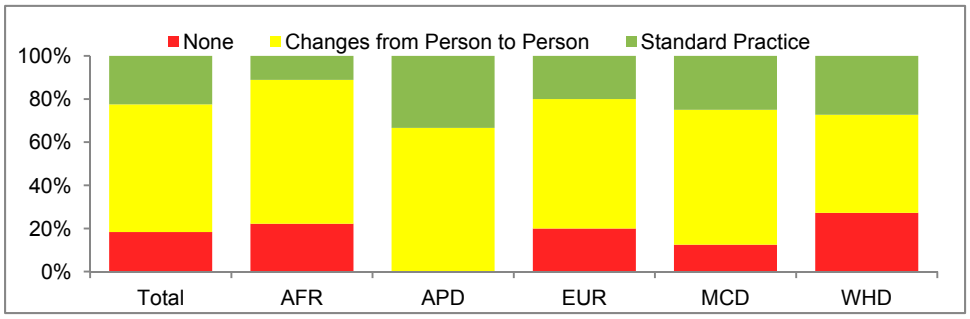
104. Similar features are found for desk economists who have worked on a greater number of country desks. In particular, the more countries a desk economist has worked on, the more important judgment becomes, the less important forecasts produced by the authorities become, the less important the use of models becomes, and the less important guidance from the department becomes. Thus, the more experience a desk economist has working on different economies, the more his forecasts tend to rely on his own judgment and the less on other sources of information.

105. Given how dependent desk economists are initially on the methods used by their predecessors, it is important that the transition between desk economists function smoothly. The more time a desk economist has to spend "figuring out the ropes" the less time he can spend analyzing the economy and improving the methodology. In follow-up interviews, most staff members indicated that transitions between country desks were ad hoc and varied

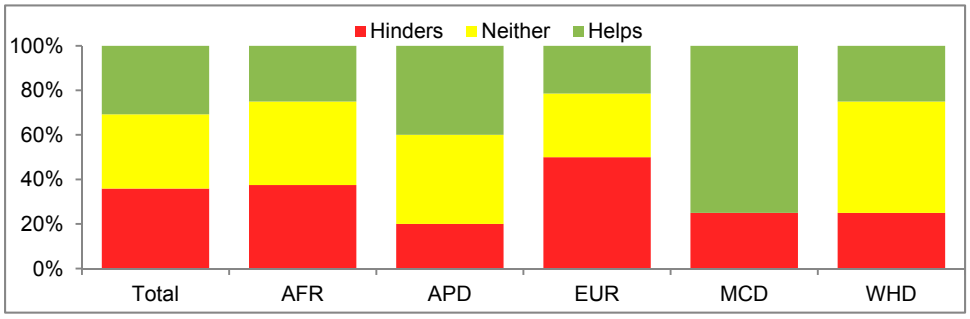
substantially from person to person (Figure 9, panel 1). While some thought the process worked fine, and several thought that the standardization of spreadsheets through DMX (Data Management for Excel)[40] had resulted in improvements, many expressed frustration with how much variation there was. In fact a view shared by several desk economists was that the only thing facilitating transitions between country desks was "good will" on the part of the outgoing desk economist.

Figure 9. Preserving memory

What mechanisms exist for transmitting accumulated knowledge of the country between the old and the new desk economist?



Does the way in which data and models are passed from the incumbent to a new desk economist help or hinder the process of generating forecasts?



Sources: Authors' calculations based on structured interviews with 50 IMF staff members.

106. The way information is transferred from outgoing to incoming economists appears to have hurt the forecasting process at the country desk level. Around 40 percent of the staff interviewed thought that the ad hoc process of transitions between country desks was hindering the process of generating the forecasts (Figure 9, Panel 2).[41] Some argued that the lack of a standard transition mechanism between country desks helped preserve the status

[40] DMX is an extension to Excel that provides tools and services to help with macroeconomic data management where data are stored in the form of time series, formulas, and tables.

[41] Figure 9 may under-represent the concerns expressed by staff: several staff members said that the way information was transferred between desks was helpful to the production of the forecasts only where the process functioned well—which is not always guaranteed.

quo and led to inertia in making changes. Others said that "a tremendous amount of information gets lost" because there is no standard way to convey this information.

(v) Conclusions and recommendations

107. The results suggest that greater staff experience is associated with smaller absolute forecast errors. We find that both country-specific and general experience help improve the forecasts.

108. However, the results are not uniform across all types of countries. While for low-income countries country-specific experience is associated with an improvement in forecast performance, for advanced and emerging economies there appears to be little relation with the forecast errors. Possibly this is because the reliance on judgment is much more prevalent in IMF forecasts for low-income countries, given the limited amount of data available and the paucity of external forecasts. Desk economists in advanced and developing economies often have an abundance of external information and forecasts that they can follow, which means there may be less need for them to rely on their own judgment.

109. The results also suggest that general work experience and training are related to an improvement in forecast accuracy. These include previous country desk assignments, tenure at the IMF, and the attendance at IMF courses related to forecasting. Interestingly, the length of tenure of mission chiefs does not appear to have a significant effect on their forecasting performance. This may be because mission chiefs already have a significant amount of general experience, so that their country-specific input to forecasts is relatively unaffected by how long they have been a mission chief for a given country.

110. The statistical results are corroborated by findings from IEO's staff survey. The survey illustrates that there is a gradual shift in how the forecasts are produced as desk economists gain more experience both on the current country assignment and from previous country assignments. When desk economists first join a country desk they rely largely on the knowledge of the previous incumbent, country authorities, and other external sources to produce their forecasts. As they gain more experience and learn about their country they tend to rely less on others and more on their own innovations and judgment when producing their forecasts.

111. A country desk economist's ability to improve the forecasts for his/her country depends on how quickly she can learn and understand the available information. An important element in this process is the transition from one country desk to the next. A poor transition can limit how quickly a desk economist is able to learn about the economy and therefore limit how much he can improve and build upon the forecasting process at his country desk. Thus, it is important that the IMF pay more attention to facilitating the transition between country desks. While some departments have begun to do this,[42] it might

[42] Several interviewees mentioned that the Fiscal Affairs Department does very well in this area.

be helpful if there were standardized best practices across the entire institution since desk economists often move across departments. Additionally, it might be valuable to allow for greater overlap between desk economists and for potential overlap during missions to encourage sharing of information between incoming and outgoing economists. While this has cost implications for the institution, the increased efficiency of incoming economists may outweigh the additional costs.

112. The results also suggest that continued importance should be given to extending and documenting desk economist tenure at the IMF. This recommendation has been a common refrain in previous reviews of the IMF's operations, and the fact that a statistically significant relationship is found between desk economist tenure and forecasting performance should add further importance to it. It seems particularly important as the IMF produces longer-term forecasts, for which greater staff experience and understanding of the structure of the economy is even more relevant. It is important to mention that the IMF has already started to heed these calls and is working to extend its average desk economist tenure to three years (IMF, 2013). While this effort will surely take some time to implement, the results of our analysis suggest that the greatest benefit in terms of forecast performance may be gained from extending the tenure of those desk economists who work in low-income countries.[43]

113. Future analysis will have to see whether there is a limit to how much experience and training can improve forecast performance; diminishing returns may set in at some point. It is also important for the institution to balance the trade-off between country-specific experience and more general experience garnered through multiple country assignments.

C. Do *WEO* Forecasts Take Account of International Linkages?

114. One of the benefits of the *WEO* forecast process is that it provides an institutionalized coordination mechanism for incorporating interrelationships between economies in individual country forecasts. As described in Genberg, Martinez and Salemi (2014), the mechanism brings together information from a global econometric model, from economic developments in major economies and regions, and from experts in fiscal and financial market analysis and transmits this information to economists responsible for the forecasts for each member country.

115. An important question in this context is whether the resulting forecasts take appropriate account of the interlinkages that exist among member countries. The answer depends on whether the relevant information from other countries and regions is effectively

[43] It is important to note that our analysis has relied on only a relatively small sample of IMF desk economist positions over a relatively short and volatile time period. Better reporting of internal data on desk economist positions on country desks could improve the Fund's ability to understand the issues and target its solutions. According to IMF (2013), "departments have just started to report desk tenure systematically; and coverage is not yet complete." Therefore, it may be several years before one can see whether these issues have been addressed.

transmitted, and whether this information is appropriately incorporated into the individual country forecasts.

116. In his review of the accuracy of IMF forecasts, Timmermann (2006) presented evidence bearing on this question. He showed that errors in growth forecasts for some countries could be explained by the forecasts of U.S. (or German) GDP growth. Put differently, forecast accuracy could have been improved if the desk economists in the identified countries had incorporated the information from the U.S. (or German) forecasts. As reported in Chapter II above, the most recent data yield results similar to those of Timmermann.

117. In this section we analyze whether IMF GDP forecasts as published in the *WEO* take appropriate account of the global factors that impinge on each economy's growth performance. To determine what is appropriate we draw on the framework presented by Matheson (2013), in which GDP growth in a country is decomposed into three factors: a global factor common to all economies; a regional factor common to the economies in the region the country belongs to; and an idiosyncratic factor that is specific to the country in question.[44] In the next section we show how this methodology can in principle be used to study whether country forecasts take appropriate account of global and regional factors, where "appropriate" is determined by what Matheson (2013) found in the empirical estimations.

118. The empirical results we present indicate that *WEO* forecasts do incorporate interdependencies among economies to a significant extent. Although we cannot answer unambiguously whether interdependencies are taken account of to an "appropriate" or optimal degree, the correspondence between the theoretical predictions of our approach and the empirical results is noteworthy. For this reason we believe that the methodology proposed here is worth pursuing further as an additional tool to assess the efficiency of *WEO* forecasts.

(i) Methodology

119. Consider the following decomposition of year-over-year real GDP growth at time *t* in country *i*, belonging to region *j* ($X_{ij,t}$):[45]

$$X_{ij,t} = A_{ij,0}G_t + A_{ij,1}G_{t-1} + B_{ij,0}R_{j,t} + B_{ij,1}R_{j,t-1} + \Psi_{ij,t} \qquad (7)$$

[44] Kose and others (2005) applied a similar structure to the G7 countries.

[45] This is the framework used in Matheson (2013).

where G, the global factor, R_j, the regional factor j, and Ψ_{ij}, the idiosyncratic factor for country ij evolve according to:

$$\begin{aligned} G_t &= CG_{t-1} + \mu_t^G \\ R_{j,t} &= D_j R_{j,t-1} + \mu_{j,t}^R \\ \Psi_{ij,t} &= E_{ij}\Psi_{ij,t-1} + \mu_{ij,t}^I \end{aligned} \tag{8}$$

and where $\mu_t^G, \mu_{j,t}^R, and\ \mu_{ij,t}^I$ are serially and mutually uncorrelated random variables for all i and j with variances equal to $\sigma^G,\ \sigma_j^R,\ and\ \sigma_{i,j}^I$ respectively.

120. A common measure of the importance of a factor in accounting for growth in a given country is the variance of the forecast error of growth accounted for by this factor. For the model described by equations (7) and (8) above this measure can be calculated quite readily (see Annex 8). The results can be given intuitive explanations.

121. The forecast error variance accounted for by the global factor *(FEV(G))* in a given country will be higher, the greater is the sensitivity of that country's growth rate to the global factor—as measured by the coefficients A in equation (7)—and the larger is the variance of shocks to the global factor (σ^G). Similar relationships exist between the forecast error variance accounted for by the regional factor (FEV(R)) and the Bs in equation (7) and the variance of the regional factor.

122. Matheson (2013) estimates the dynamic factor model represented by (7) and (8) and provides estimates of both *FEV(G)* and *FEV(R)* for 185 countries using data for 1990–2011. We take these values as benchmarks for interdependence among these countries and ask whether *WEO* forecasts incorporate this interdependence.

123. A natural way to do this would be to estimate a model of the same form as that used by Matheson but to use *WEO* forecasts, rather than actual values, of growth rates as the dependent variable. Unfortunately, for the majority of IMF members this approach is made difficult by the fact that *WEO* forecasts are only available semi-annually, while estimation of the dynamic factor model requires more frequent observations. For this reason we converted Matheson's dynamic factor model into a static model by working with forecast *revisions* rather than the forecasts themselves. From this model we can calculate the fraction of the variance of forecast revisions *(FRV(G))* that is accounted for by the global factor.[46]

124. Theoretically the forecast-error variance decompositions and the forecast-revision decomposition are related, as they are all determined by the parameters defining equations (7) and (8). Specifically it can be shown that:

- A country for which *FEV(G)* is high (low) should also have a high (low) *FRV(G)*.

[46] See Annex 8 for details.

- A country for which *FEV(R)* is high may have a high or low *FRV(G)* depending on the source of the high *FER(R)*, i.e., whether it is due to a high value of σ_R relative to σ_I or a high value of *B*.

- If the value of *FEV* increases over time—either because *A* increases relative to *B* or because the variance of the global shocks increases relative to the regional or idiosyncratic shocks—then *FRV* should increase over time as well.

125. We estimate *FRV(G)* and investigate whether the relationships between this measure and the forecast error variances estimated by Matheson conform to the theoretical predictions derived above.

(ii) Data, estimation method, and results

Data and estimation method

126. To estimate the effect of a global factor on *WEO* forecast revisions we collected data on *WEO* forecast revisions for 143 countries for the period 1991 to 2011 (see Annex 9 for the list of countries). We used two different revisions per year per country in the analysis: the revision of the forecast of year *t+1* between the Fall of year *t* and the Spring of year *t+1*, and the revision of the forecast of year *t+1* between the Spring of year *t+1* and the Fall of year *t+1*.

127. Using these data we estimate a model that corresponds to the theoretical representation in equation (7) above where the regional and idiosyncratic terms are combined into one. In other words, we estimate a model where the forecast revisions for each country in the sample are influenced by a common latent global factor and a country-specific idiosyncratic factor, as in equation (9):

$$FR_{i,t} = \alpha_i G_t + \varepsilon_{i,t} \qquad (9)$$

128. The factor loadings for each country (denoted by α) represent the sensitivity of the forecast revision to the global factor. Based on this relationship the estimated contribution of the global factor to the variance of the forecast revision for country *i* is simply:

$$FRV(G)_i = \frac{\alpha_i^2}{\alpha_i^2 + \sigma_{\varepsilon_i}^2} \qquad (10)$$

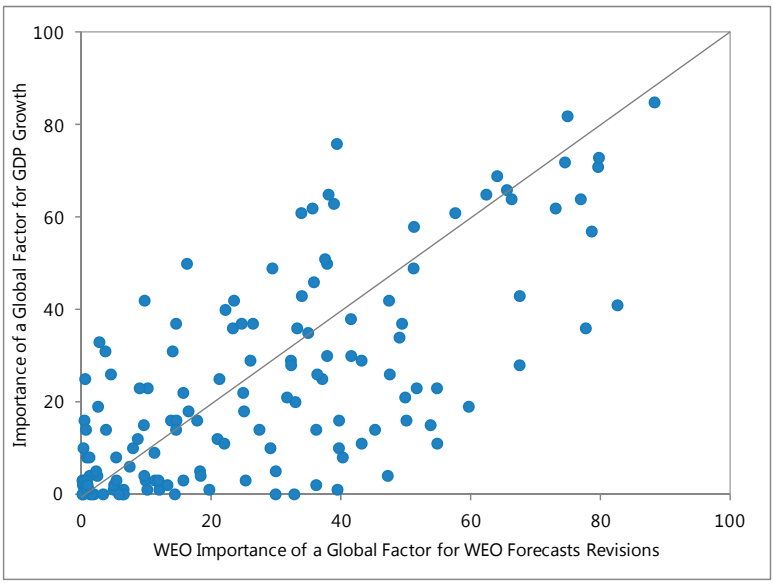
where the variance of the latent global factor *G* has been normalized to 1. Estimation of the model was carried out in STATA using the structural equation model command *(sem)*.

Results

129. A first view of the results is given in Figure 10, whose vertical axis shows the forecast error variances attributable to the global factor in the study by Matheson and whose

horizontal axis shows the forecast revision variances attributable to the global factor as obtained from our estimation. Recall from the discussion above that there should be a positive relationship between these two variables if (i) Matheson's model is a good representation of interdependencies among the IMF member economies and (ii) IMF country desk economists incorporate these interdependencies in their forecasts. The clear positive association between the two variables is a sign that the *WEO* forecasts produced by country desk economists incorporate some of the patterns of interdependencies characterizing actual data on GDP growth.

Figure 10. Comparing "actual" and estimated interdependence



Source: Authors' calculations using *IMF WEO* and Matheson (2013).

130. While the theoretical discussion showed that the forecast error variances and forecast-revision variances are related, it also showed that the relationship between them depends on the nature of the economies involved, as described in part by the importance of regional effects and in part by the nature of the relationship in the data between the economy and the global and regional factor, the *A*s and *B*s in equation (7).

131. In an attempt to capture these more subtle effects we estimate a model where the relationship between *FEV* and *FRV* depends on the type of economy that is considered, specifically whether it is an advanced *(ADV)*, emerging *(EME)*, or low-income country *(LIC)* economy, as described in equations (11), (11a), and (11b).

$$FRV_i = \beta_0 + \beta_1 FEV(G)_i + \beta_2 FEV(R)_i + v_i \quad (11)$$

$$\beta_0 = \beta_{00} + \beta_{0,EME} EME_i + \beta_{0,LIC} LIC_i \quad (11a)$$

$$\beta_1 = \beta_{1,ADV} ADV_i + \beta_{1,EME} EME_i + \beta_{1,LIC} LIC_i \quad (11b)$$

132. If *WEO* forecasts incorporate interdependencies in the data (as measured by Matheson) we would expect β_1 to be positive. Furthermore, we expect $\beta_{1,ADV}$ to be larger than $\beta_{1,EME}$ and $\beta_{1,LIC}$ as a result of the generally greater integration of advanced economies in the global economy. For the same reason we also expect $\beta_{0,0}$ to be larger than $\beta_{0,EME}$ and $\beta_{0,LIC}$. Finally, based on the arguments in the paragraphs following equation (9), the sign of β_2 could be either positive or negative depending on the sources of the importance of the regional factor.

133. As shown in Table 13, the estimation results correspond quite closely with prior expectations, and the explanatory power of the model, at 62 percent, is substantial given that *WEO* forecasts are generated independently by each country desk economist. We view this at least in part as being due to the coordination built into the *WEO* process.

Table 13. Estimates of results based on equation (11)

	Estimated FRV(G)
Constant	42.648 (8.604)***
Emerging Markets	-11.804 (8.383)
LICs	-20.287 (8.109)**
Advanced * Matheson's FEV(G)	0.431 (0.131)***
Emerging * Matheson's FEV(G)	0.257 (0.107)**
LIC * Matheson's FEV(G)	0.435 (0.138)***
Matheson's FEV(R)	-0.183 (0.064)***
R^2	0.62
N	143
* $p<0.1$; ** $p<0.05$; *** $p<0.01$	

Source: Authors' calculations.

134. While we view these results as reflecting favorably on the *WEO* forecast process, we cannot determine whether *WEO* forecasts reflect interdependencies among economies optimally.[47] This is because we have been constrained to estimate a static factor model for forecast revisions as opposed to the "true" dynamic model for the forecasts themselves.

[47] By "optimally" we mean "as measured by Matheson." The statement is conditional on the model used by Matheson being the correct one.

(iii) Comparing with forecasts from Consensus Economics[48]

135. Not able to determine whether *WEO* forecasts incorporate interdependencies optimally we compared the revisions in *WEO* forecasts with revisions in those provided by Consensus Economics, using the forecast revision variance as a metric. Only for a subset of countries do consensus forecasts exist on a consistent basis from 1990–2011. Thus, the comparison could be made for 21 countries only (see Annex 8 for details).[49]

136. In a first comparison we estimated the same regression model as in (11) except that β_0 and β_1 did not depend on country groupings.[50] The results, presented in Table 14, show that Consensus forecasts are more sensitive to the "true" interdependencies reflected in Matheson's variance decompositions—in the sense both that the coefficient on Matheson's *FEV(G)* is larger and that the R^2 is higher. Does this mean that Consensus forecasts capture the true interdependencies better? Unfortunately this is another question we cannot answer unequivocally, because of the lack of one-to-one theoretical correspondence between *FEV(G)* and *FER(G)*. But the results strike us as somewhat surprising since *WEO* forecasts typically contain a more elaborate top-down element than Consensus forecasts (see Genberg, Martinez and Salemi, 2014), and would therefore be more likely to reflect global factors.

Table 14. Comparing Consensus economics and *WEO* forecasts

	Consensus	*WEO*
Matheson's FEV(G)	0.510 (0.174)***	0.407 (0.204)*
Mathesons's FEV(R)	-0.206 (0.213)	-0.184 (0.250)
Constant	51.401 (14.785)***	50.339 (17.381)***
R^2	0.53	0.35
N	21	21
* $p<0.1$; ** $p<0.05$; *** $p<0.01$		

Source: Authors' calculations.

137. In a further exploration of the relationship between *WEO* and Consensus forecasts we estimated the factor model for sub-periods to see whether we could detect any changes over time in the role played by the global factor in the respective forecasts.[51] Figure 11 tells an

[48] Isiklar and others (2006) also investigate whether Consensus forecasts take into account international linkages and finds that they do not do so fully.

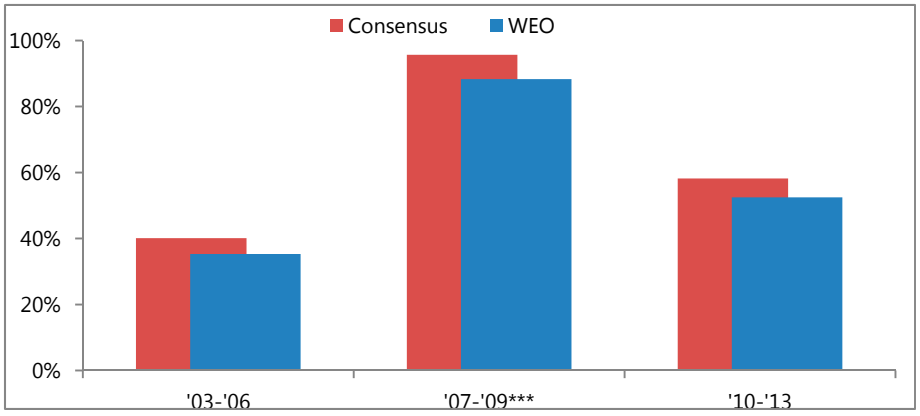
[49] For the purpose of the comparison we re-estimated the model described by equation (10) for the common sample of countries.

[50] Consensus Economics does not provide forecasts for LICs and we considered the number of EMEs too small to provide meaningful estimates for our purposes.

[51] Since the estimation only goes back to 2003, a larger sample of 72 countries could be used. See Annex 9.

interesting story. It shows the simple average across countries in the *FRV* attributable to the global factor. Three features of the results are noteworthy. First, the average *FRV(G)* is consistently higher for the Consensus than the *WEO* forecasts, although the difference is only statistically different from zero in 2007–09. This is consistent with the results reported in Table 14. Second, there is a substantial increase in the role of the global factor during the 2007–09 "crisis years" for both forecasters, suggesting a higher variance of the global factor during these years compared with pre- and post-crisis years. Third, the role of the global factor is larger in the 2010–13 period than in the 2003–06 period, consistent with the idea that the world economy has become more interdependent over time, as measured by the size of the coefficients A_0 and A_1 in equation (7).

Figure 11. Estimated contributions of the global factor to forecast revisions[1]



Source: Authors' calculations using *IMF WEO* and Consensus Economics.

[1]Asterixes denote years when differences between *WEO* and Consensus are statistically significant based on a t-test for equality of means: *p<0.10; **p<0.05; *** p<0.01.

(iv) Conclusions

138. We have found that *WEO* forecasts do incorporate interdependencies among economies to a substantial degree. Although we have not been able to answer unambiguously whether interdependencies are incorporated to an "appropriate" or optimal degree, the correspondence between the theoretical predictions of our approach and the empirical results is noteworthy. For this reason we believe that the methodology proposed here is worth pursuing further as an additional tool to assess the efficiency of *WEO* forecasts.

IV. LESSONS

139. Rather than repeating specific conclusions that have been presented in individual sections we focus on three broad lessons that can be drawn from the analysis.

A. Biases in *WEO* Forecasts

140. Statements of the type "*WEO* growth forecasts are typically too optimistic" or "*WEO* forecasts of GDP growth are generally pessimistic" are too simplistic and do not find support in the data. Though it is possible to identify periods where *WEO* growth forecasts on average across the membership have been higher than out-turns, it is also possible to find periods when the opposite has been the case. Furthermore, even in periods where on average an optimistic bias was present, it is often possible to identify some subgroups of countries for which forecasts have erred on the pessimistic side.

141. Finding that *WEO* forecasts are not consistently biased in one direction or the other should not be a reason for complacency. Lack of bias only means that positive and negative forecast errors tend to cancel each other out over time. It does not mean that forecast errors are small or that there are no possibilities for improvement. Indeed there are periods where forecast errors have been substantial and relatively widespread. A pertinent question therefore is whether it is possible to identify institutional reasons why in some time periods the forecasts for a large number of member countries tend to err in the same direction.

142. For the membership as a whole, large forecast errors tend to be particularly clustered around regional or global recessions. Might there be institutional reasons for this? While it is clear that some events may be unpredictable, Juhn and Loungani (2002) have argued that the failure of private sector forecasters to predict recessions could arise because they lack the incentives to do so.

143. Because the internal forecasting process at the IMF does not encourage forecasts that "rock the boat," IMF staff too may lack incentives to predict recessions. As part of the review process, staff forecasts are checked against those of other forecasters and need to be justified if they are different. Thus, desk economists can minimize the amount of scrutiny their forecasts will get by not differing significantly from mainstream opinion. While this scrutiny operates symmetrically, the cost of forecasting a recession that does not materialize may be perceived as higher than that of having wrongly predicted a boom.

B. Are Interdependencies Among Economies Accounted for?

144. Second, while we have found evidence that *WEO* forecasts incorporate interdependence among countries to a significant degree, linkages between economies may still not be fully incorporated in all forecasts. In his comprehensive 2006 study Timmermann pointed out that using information contained in forecasts of the U.S. and German economies could have reduced the *WEO*'s forecast errors for a number of other countries. Using updated

data we find that this is still the case. In addition, we find that forecasts of the Chinese economy appear to contain information that could have improved other forecasts.

145. The *WEO* forecast process contains elements that are designed to increase individual desk economists' awareness of relevant international developments (Genberg, Martinez and Salemi, 2014). The evidence just referred to suggests that these elements may need to be strengthened.

C. Can More Be Learned From Past Experience?

146. The third broad lesson relates to learning. Longer experience of desk economists appears to result in smaller forecast errors, especially in the case of low-income countries. Experience can be gained in several ways; through learning-by-doing as a desk economist for a given country or in successive assignments in countries where similar forecasting challenges are present; by receiving guidance and information from a predecessor at a country desk; and by formal training in courses designed to enhance forecasting skills. Each of these sources of experience has clear implications for policies and practices relating to the nature of country assignments, to the process of moving staff between country assignments, and the organization of relevant training courses for staff and incentives of staff to avail themselves of such courses.

147. Learning can also be seen in a broader perspective. The IMF could devote resources to providing guidance on best practice in forecasting in countries with particular characteristics—such as commodity-rich countries; data-poor environments; highly export-dependent countries; countries with fixed exchange rates; etc. Systematically studying past successes and failures in forecasting in different environments might provide valuable lessons that desk economists can draw on when they face the challenges of a new country assignment.

REFERENCES

Abreu, Ildeberta, 2011, "International Organizations' vs. Private Analysts' Growth Forecasts: An Evaluation," *Banco de Portugal Economic Bulletin*, Vol. 17 (Summer), pp.23–44.

Aldenhoff, Frank-Oliver, 2007, "Are economic forecasts of the International Monetary Fund politically biased? A public choice analysis," *Review of International Organizations*, Vol. 2 (September), pp.239–260.

Arora, Harji K., and David J. Smyth, 1990, "Forecasting the developing world: An accuracy analysis of the IMF's forecasts," *International Journal of Forecasting,* Vol. 6 (October), pp.393–400.

Artis, Michael 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*, pp.1–49, July **(**Washington: International Monetary Fund).

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

_____, 1997, "How Accurate are the IMF's Short-Term Forecasts? Another Examination of the *World Economic Outlook*," *Staff Studies for the World Economic Outlook*, pp.1–39, December **(**Washington: International Monetary Fund).

Artis, Michael J., and Wenda Zhang, 1990, "BVAR forecasts for the G-7," *International Journal of Forecasting,* Vol. 6 (October), pp.349–362.

Artis, Michael J., and Massimiliano Marcellino, 2001, "Fiscal forecasting: The track record of the IMF, OECD, and EC," *Econometrics Journal*, Vol. 4 (June), pp.S20–S36.

Ashiya, Masahiro, 2006, "Testing the Rationality of Forecast Revisions Made by the IMF and the OECD," *Journal of Forecasting*, Vol. 25 (January), pp.25–36.

Atoyan, Rouben, Patrick Conway, Marcelo Selowsky, and Tsidi Tsikata, 2004, "Macroeconomic Adjustments in IMF-Supported Programs: Projections and Reality," IEO Background Paper BP/04/2, April (Washington: International Monetary Fund). Available at: www.ieo-imf.org.

Atoyan, Ruben, and Patrick Conway, 2011, "Projecting macroeconomic outcomes: Evidence from the IMF," *Review of International Organizations*, Vol. 6 (September), pp.415–441.

Baba, Chikako, and Turgu Kisinbay, 2011, "Predicting recessions: A new approach for identifying leading indicators and forecast combinations," IMF Working Paper No. WP/11/235, October (Washington: International Monetary Fund).

Baqir, Reza, Rodney Ramcharan, and Ratna Sahay, 2005, "IMF Programs and Growth: Is Optimism Defensible?" *IMF Staff Papers*, Vol. 52 (September), pp.260–286.

Batchelor, Roy, 2001, "How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the consensus," *Applied Economics*, Vol. 33 (February), pp.225–235.

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

Beach, William, Aaron Schavey, and Isabel Isidro, 1999, "How Reliable Are IMF Economic Forecasts?" Heritage Foundation Center for Data Analysis Report 99-05, August (Washington: Heritage Foundation).

Blanchard, Olivier, and Daniel Leigh, 2013, "Growth Forecast Errors and Fiscal Multipliers," IMF Working Paper WP/13/1, January (Washington: International Monetary Fund).

Blix, Mårten, Joachim Wadefjord, Ulrike Wienecke, and Martin Ådahl, 2001, "How good is the forecasting performance of major institutions?" *Sveriges Riksbank Economic Review*, March (Stockholm: Sveriges Riksbank).

Boettcher, Michaela, 2004, "Modeling and forecasting activities of the International Monetary Fund," *mimeo*.

Cabanillas, Laura Gonzalez, and Alessio Terzi, 2012, "The accuracy of the European Commission's forecasts re-examined," European Commission, European Economy Economic Papers, No. 476, December (Brussels: European Commission).

Clement, Michael B., 1999, "Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter?" *Journal of Accounting & Economics*, Vol. 27 (July), pp.285–303.

Clement, Michael B., Lyn Rees, and Edward P. Swanson, 2003, "The Influence of Culture and Corporate Governance on the Characteristics That Distinguish Superior Analysts," *Journal of Accounting, Auditing & Finance*, Vol. 18 (October), pp.593–618.

De Masi, Paula R., 1996, "The Difficult Art of Forecasting," *Finance & Development*, Vol. 33 (December), pp.29–31.

Diebold, F.X. and Lopez, J.A. (1996), "Forecast Evaluation and Combination," in G.S. Maddala and C.R. Rao (eds.), *Handbook of Statistics*. Amsterdam: North-Holland, 241–268.

Doepke, Joerg, and Ulrich Fritsche, 2004, "Growth and Inflation Forecasts in Germany—An Assessment of Accuracy and Dispersion," DIW Discussion Papers 309, February (Berlin: German Institute for Economic Research).

Drechsel, Katja, Sebastian Giesen, and Axel Lindner, 2013, "Outperforming IMF Forecasts by the Use of Leading Indicators," February, *mimeo.*

Dreher, Axel, Silvia Marchesi, and James Raymond Vreeland, 2008, "The political economy of IMF forecasts," *Public Choice*, Vol. 137 (October), pp.145–171.

Elliott, Graham, Ivana Komunjer, and Allan Timmermann, 2005, "Estimation and Testing of Forecast Rationality under Flexible Loss," *Review of Economic Studies*, Vol. 72 (October), pp.1107–1125.

Ericsson, Neil R., 2013, "How Biased Are U.S. Government Forecasts of the Federal Debt?" June, *mimeo*.

European Central Bank, 2012, "Forecast Bias for Euro Area HICP Inflation," ECB Monthly Bulletin, June (Frankfurt: European Central Bank).

Faust, Jon, 2013, "A report on the predictive accuracy of the IMF's *WEO* forecast," *mimeo.*

Fratianni, Michele, and John Pattison, 1991, "International Institutions and the Market for Information," in Roland Vaubel and Thomas D. Willett, eds., The Political Economy of International Organizations: A Public Choice Approach (Boulder: Westview Press).

Freedman, Charles, 2014, "An Evaluation of Commissioned Studies Assessing the Accuracy of IMF Forecasts," IEO Background Paper BP/14/02 (Washington: Independent Evaluation Office of the International Monetary Fund).

Frenkel, Michael, Jan-Christoph Ruelke, and Lilli Zimmermann, 2013, "Do private sector forecasters chase after IMF or OECD forecasts?" *Journal of Macroeconomics*, Vol. 37 (April) pp.217–229.

Garcia-Meca, Emma, and Juan Pedro Sanchez-Ballesta, 2006, "Influences on financial analyst forecast errors: A meta-analysis," *International Business Review*, Vol. 15 (February), pp.29–52.

Genberg, Hans, and Andrew Martinez, 2014, "User Perspectives on IMF Forecasts: Survey Methodology and Results," IEO Background Document No. BD/14/01 (Washington: Independent Evaluation Office of the International Monetary Fund).

Genberg, Hans, Andrew Martinez, and Michael Salemi, 2014, "The IMF/*WEO* Forecast Process," IEO Background Paper BP/14/03 (Washington: Independent Evaluation Office of the International Monetary Fund).

Glück, Heinz, and Stefan Schleicher, 2005, "Common Biases in OECD and IMF Forecasts: Who Dares to be Different?" *mimeo*.

Golosov, Mikhail, and John King, 2002, "Tax Revenue Forecasts in IMF-Supported Programs," IMF Working Paper WP/02/236, December (Washington: International Monetary Fund).

Hawkins, John, 2002, "Cassandra and the Sirens: Economic forecasting in emerging economies," *IFC Bulletin* 13, December (Washington: IFC).

Hendry, David F., and Carlos Santos, 2010, "An Automatic Test of Super Exogeneity," in T. Bollerslev, J.R. Russell, and M.W. Watson, eds., *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle* (Oxford: Oxford University Press).

Hendry, David F., Soren Johansen, and Carlos Santos, 2008, "Automatic selection of indicators in a fully saturated regression," *Computational Statistics*, Vol. 23 (April), pp.317–335.

Independent Evaluation Office of the International Monetary Fund (IEO), 2003, *The IMF and Recent Capital Account Crises* (Washington: IEO). Available at: www.ieo-imf.org.

_____, 2006, "An Evaluation of the IMF's Multilateral Surveillance—Background Documents," (Washington: IEO). Available at: www.ieo-imf.org.

_____, 2009, *IMF Interactions with Member Countries* (Washington: IEO). Available at: www.ieo-imf.org.

_____, 2011, *IMF Performance in the Run-Up to the Financial and Economic Crisis* (Washington: IEO). Available at: www.ieo-imf.org.

_____, 2013, *The Role of the IMF as Trusted Advisor* (Washington: IEO). Available at: www.ieo-imf.org.

International Monetary Fund, 2013, "2013 Corporate Workforce Planning," FO/Dis/13/24, February (Washington).

Isiklar, Gultekin, Kajal Lahiri, and Prakash Loungani, 2006, "How quickly do forecasters incorporate news? Evidence from cross-country surveys," *Journal of Applied Econometrics*, Vol. 21 (October), pp.703–725.

Jacob, John, Thomas Lys, and Margaret Neale, 1999, "Expertise in forecasting performance of security analysts," *Journal of Accounting & Economics*, Vol. 28 (November), pp.51–82.

Jagric, Timotej, and Jani Beko, 2011, "How Good Are the Growth and Inflation Forecasts for Slovenia?" *Romanian Journal of Economic Forecasting*, Vol. 14 (December), pp.47–67.

Johansen, Soren, and Bent Nielsen, 2009, "An Analysis of the Indicator Saturation Estimator as a Robust Regression Estimator," in J.L. Castle and N. Shepard, eds., *The Methodology and Practice of Econometrics: A Festschrift in Honour of David Hendry* (Oxford: Oxford University Press).

Juhn, Grace, and Prakash Loungani, 2002, "Further Cross-Country Evidence on the Accuracy of the Private Sector's Output Forecasts," *IMF Staff Papers*, Vol. 49 (April), pp.49–64.

Julio, Paulo, and Pedro Esperanca, 2012, "Evaluating the forecast quality of GDP components: An application to G7," GEE Papers 47 (Lisboa: Ministerio da Economia E do Emprego).

Kenen, Peter B., and Stephen B. Schwartz, 1986, "An Assessment of Macroeconomic Forecasts in the International Monetary Fund's *World Economic Outlook*," Working Paper in International Economics G-86-04 (Princeton: Princeton University).

Kose, M. Ayhan, Christopher Otrok, and Charles H. Whiteman, 2005, "Understanding the Evolution of World Business Cycles," IMF Working Paper WP/05/211, November (Washington: International Monetary Fund).

Kreinin, Mordechai E., 2000, "Accuracy of OECD and IMF Projection," *Journal of Policy Modeling*, Vol. 22 (January), pp.61–79.

Krkoska, Libor, and Utku Teksoz, 2009, "How reliable are forecasts of GDP growth and inflation for countries with limited coverage?" *Economic Systems*, Vol. 33 (xxx), pp 376–388.

Loungani, Prakash, 2001, "How accurate are private sector forecasts? Cross-country evidence from *consensus forecasts* of output growth," *International Journal of Forecasting*, Vol. 17 (September), pp.419–432.

Luna, F., 2014a, " IMF Training on Forecasting and Forecasting Method, " IEO Background Document No. BD/14/02 (Washington: Independent Evaluation Office of the IMF).

Luna, F., 2014b, "Forecasts in the Context of Program Countries," IEO Background Paper No. BP/14/05 (Washington: Independent Evaluation Office of the IMF).

Matheson, Troy, 2013, "The Global Financial Crisis: An Anatomy of Global Growth," IMF Working Paper WP/13/76, March (Washington: International Monetary Fund).

Melander, A., G. Sismanidi, and D. Gernouilleau, 2007, "The track record of the Commission's forecasts—an update," European Commission, European Economy Economic Papers, No. 291, October (Brussels: European Commission).

Mikhail, Michael B., Beverly R. Walther, and Richard H. Willis, 1997, "Do Security Analysts Improve Their Performance with Experience?" *Journal of Accounting Research*, Vol. 35, pp.131–157.

Mrkaic, Mico, 2010, "Data Dissemination Standards and the Statistical Quality of the IMF's World Economic Outlook Forecasts," IMF Working Paper WP/10/203, September (Washington: International Monetary Fund).

Musso, Alberto, and Steven Phillips, 2002, "Comparing Projections and Outcomes of IMF-Supported Programs," *IMF Staff Papers*, Vol. 49 (April), pp.22–48.

Pisani-Ferry, Jean, Andre Sapir, and Guntram Wolff, 2011, "An evaluation of IMF surveillance of the euro area," *Bruegel Bluepring* 14.

Pons, Jordi, 2000, "The Accuracy of IMF and OECD Forecasts for G7 Countries," *Journal of Forecasting*, Vol. 19 (January), pp.53–63.

Takagi, Shinji, and Halim Kucur, 2006, "Testing the Accuracy of IMF Macroeconomic Forecasts, 1994–2003," IEO Background Paper No. BP/06/1, May (Washington: International Monetary Fund). Available at: http://www.ieo-imf.org.

_____, 2008, "Comparing the Information Contents of IMF and OECD Macroeconomic Forecasts," *Osaka Economic Papers*, Vol. 58 (September), pp.269–282.

Timmermann, Allan, 2006, "An Evaluation of the World Economic Outlook Forecasts," IMF Working Paper WP/06/59, March (Washington: International Monetary Fund).

_____, 2007, "An Evaluation of the *World Economic Outlook* Forecasts," *IMF Staff Papers*, Vol. 54 (June), pp.1–33.

Tong, Hui, 2004, "Do Transparency Standards Improve Macroeconomic Forecasting?" *mimeo*.

U.S. Government Accountability Office (GAO), 2003, "International Financial Crises: Challenges Remain in IMF's Ability to Prevent and Resolve Financial Crises," GAO-03-734, June (Washington: Government Accountability Office).

Vaubel, Roland, 2009, "Lessons from the Financial Crisis: The International Dimension," *Economic Affairs*, Vol. 29 (September), pp.22–26.

Verbeek, Jos, 1998, "The World Bank's Unified Survey Projections, How Accurate Are They? An Ex-Post Evaluation of US91-US97," World Bank Policy Research Working Paper 2071, December (Washington: World Bank).

West, Kenneth, (2006), "Forecast Evaluation," 100-134 in *Handbook of Economic Forecasting*, Vol. 1, G. Elliott, C. Granger and A. Timmerman (eds), Amsterdam: Elsevier.

Zeng, Li, 2011, "Evaluating GDP Forecasting Models for Korea," IMF Working Paper WP/11/53, March (Washington: International Monetary Fund).

ANNEX 1. DATA ON IMF FORECASTS

IMF forecasts are produced by and available through several different sources. The primary source for these forecasts is the *World Economic Outlook* report, published twice a year, and its corresponding database. IMF forecasts are also produced in the course of the Fund's regular Article IV consultations as well as in program documents for each country with a Fund program, but the irregular frequency of these consultations makes it hard to evaluate the forecasts' performance and means that they are not always aligned with the *WEO* forecasts. The Monitoring of Fund Arrangements (MONA) database specifically keeps track of the projections made for IMF programs. Most recently, the *Fiscal Monitor* has started maintaining a database on projections of various fiscal variables. While data exist for each of these sources, our analysis in this paper primarily refers to the *WEO* forecasts.

Accessibility of the Forecasts and the Actuals

While there is no single database for the IMF's forecasts, the recent forecasts are easily accessible through the IMF website. Each individual release of the *WEO*'s forecasts is available for more than a decade dating back through 1998, either in the statistical appendixes of the relevant *WEO* publication or in its corresponding database. Similarly, for the *Fiscal Monitor*, since 2010 a database with the forecasts has been released along with each publication.

It is much harder to access historical forecasts from the *WEO*. Despite the fact that the *WEO* has produced forecasts since 1971, and has published them since 1980, the IMF website provides no information on the forecasts prior to the late 1990s. However, these forecasts can be found in historical *WEO* publications since 1984 online through the IMF eLibrary, or since 1980 in hard copies of the publications. Additionally, several unofficial sources such as Artis (1996) and Timmermann (2007) have compiled partial historical databases of *WEO* forecasts.[1,2]

The actual values of the variables forecast by the *WEO* are available along with the forecasts. However, the data often go through several revisions such that each new database has a slightly different vintage of the data. These constant revisions of data somewhat complicate its use for evaluating the forecasts themselves. The "first available" actualized values for a given year are available in the Spring of the following year while the "first settled" values are released in the Fall of that year. After these initial releases the data are regularly updated in each new *WEO* database to reflect changes in estimates and revisions. The availability of

[1] Artis (1996) has tables on various IMF forecasts from 1971–94 while Timmermann (2007) has GDP and inflation forecasts for individual countries from 1990-2004 in a data appendix: www.imf.org/External/Pubs/FT/staffp/2007/01/timmerma.htm

[2] A preliminary review of the *WEO's* online data forum illustrates that there is at least some demand for a more complete historical database of IMF forecasts and actual data by *WEO* data users.

these different vintages of data has led to differences—sometimes consequential—among what different studies use as actual values against which to compare *WEO* forecasts.

Variable and Country Coverage

Over the past four decades the number of countries and variables for which the IMF has produced forecasts has grown substantially. Initially, the *WEO* only produced forecasts for individual G-7 countries and for groups of advanced economies. By the 1980s, however, the *WEO* was producing individual forecasts for the G-7 countries as well as forecasts for different regions of the world. It was not until 1999, with the release of its online database, that the *WEO* started publishing forecasts for all countries.[3] The *Fiscal Monitor* only publishes forecasts for a select set of countries such as G-7 countries, the Euro Area, and selected emerging markets. Furthermore, the *Fiscal Monitor*'s country coverage tends to differ with each variable.

While the *WEO's* country coverage has expanded over the past four decades, the variables for which it provides forecasts have remained fairly constant. Table A1 provides an overview of when the main *WEO* variables were first forecast for different country groups. The table illustrates that while most of the main variables in the *WEO* forecast have been produced since the beginning, several variables such as the output gap, population, and government debt are relatively new additions to the *WEO* database.

There is increasing overlap between the forecasts provided by the *WEO* and the *Fiscal Monitor*. While the *WEO* has forecast the fiscal balance for G-7 countries since the mid-1970s, its continued expansion of other fiscal variables including gross and net government debt, as well as providing these forecasts for most of the IMF membership in 2010, coincides with the introduction of the *Fiscal Monitor* and the growing importance of these variables in political discourse. Given the possibilities of differences between these databases it might make sense to unify the forecasts into a single database or to create a database that only forecasts fiscal variables. This would reduce duplication and ensure consistency in the forecasts of fiscal variables, and might help reduce confusion for forecast users.

[3] Timmermann (2007) illustrates that even though they were not published, *WEO* forecasts for GDP and inflation have been produced for all countries since at least 1990.

Table A1. Starting year for published forecasts of selected variables in the *WEO*[1]

Variable	Advanced			Emerging & LIC		World
	G-7	Groups	Countries	Regions	Countries	
National Accounts						
Real GNP / GDP	1971	1972	1995[1]	1977	1999[1]	1985
Components of GDP	1974	1982	-	-	-	-
GDP Deflator	1971	1972	1993	-	2003[2]	-
CPI	1974	1975	1995[1]	1978	1999[1]	-
Employment	1974	1985	2007	-	-	-
Unemployment rate	1976	1985	1995	-	2010	-
Interest rate	1985[3]	-	-	-	-	-
Commodity Prices	-	-	-	-	-	1975
Hourly Earnings[4]	1974	1985	-	-	-	-
Productivity[4]	1975	1985	-	-	-	-
Unit Labor Costs[4]	1974	1985	-	-	-	-
Output Gap	1993	-	2004	-	-	-
Population	2006	-	2006	-	2006	-
External						
Trade Balance	1971	1971	2011	1972	2011	1972
Exports	1971	1971	2011	1972	2011	1972
Imports	1971	1971	2011	1972	2011	1972
Current Account	1972	1972	1995	1977	2004	1973
External Debt	-	-	-	1978	-	-
Reserves	-	-	-	1981	-	-
Government Finances						
Fiscal Balance	1975	1993	1995	1986	2010	-
Gross Debt	1994	-	2010	-	2010	-
Net Debt	1994	-	2010	-	2010	-

Source: *WEO* reports since 1971 and *WEO* databases since 1999.

[1]Available starting in 1990 from Timmerman (2006) data.
[2]Can be calculated from forecasts in *WEO* database starting in 1999.
[3]LIBOR rate in US through 1990. Starting in 1991 also includes short-term rates for U.S., Japan, and Germany.
[4]Discontinued in 2007.

ANNEX 2. LITERATURE RELATED TO THE QUALITY OF IMF FORECASTS

Study	Variables	Countries	Years	Forecasters[1]	Horizon	Actuals	Tests and Stats[2]	Summary of Main Findings (related to IMF forecasts)
Worswick (1983)	GDP, inflation	n/a	'81-'82	IMF	T, T+1	n/a	n/a	WEO has an optimism bias.
Kenen and Schwartz (1986)	inflation, GDP, current account, exports, imports	21 countries	'71-'85	IMF	T, T+1	first (WEO)	reg (B, EF), PR, U	Under predicted inflation more than over predicted real growth.
Artis (1987) / Artis (1988)	GDP deflator, CPI, current account, GDP, exports, imports, terms of trade	14 countries	'71-'85	IMF, OECD, VN	T, T+1	first (WEO)	reg (B, EF), MAE, RMSFE, U	The results suggest a degree of output optimism in WEO forecasts.
Arora and Smith (1990)	GDP, exports, imports, inflation, debt, current account, reserves	6 country groups	'80-'88	IMF	T	latest (WEO)	MAE, RMSFE, MEDAE, U	IMF forecasts are inferior to a random walk model.
Artis and Zhang (1990)	output, inflation, current account	7 countries	'80-'87	IMF, Model	T, T+1	first (WEO)	U	BVAR Models perform on par with WEO forecasts.
Fratinni and Pattinson (1991)	GDP, inflation, current account	7 countries	'80-'87	IMF, OECD	T	unclear	MAE, RMSFE	There is no evidence that the published forecasts of the IOs are superior to the national forecasts.
Spring WEO (1992) / Fall WEO (1992) / Barrionuevo (1993)	GDP, inflation	14 countries, 4 country groups, 36 program countries	'71-'91	IMF, Model	T, T+1	unclear	MFE, MAE, RMSFE, U, reg (B, EF)	Updated results suggest that the bias in the WEO forecasts was reduced after 1985 Forecasts mostly superior to random walk forecasts for industrial countries, random walk superior for all developing countries.
Artis (1996) / Artis (1997) / De Masi (1996) / Fall WEO (1996)	GDP, inflation, current account, imports and exports	14 countries, 5 country groups	'71-'94	IMF, CONS	T, T+1	first (WEO)	reg (B, EF), U, ENCOMP, B, MAE, RMSFE, DA	Developing country forecasts are distinctly weaker than those for the developed industrial group. Major forecasting errors are substantially the same when compared with Consensus forecasts.
Verbeek (1998)	imports, GDP deflator, domestic investment, fiscal balance, current account, GDP, exports	23 countries	'91-'97	WB, IMF	T, T+1	latest	MAE, RMSFE, U, DECOMP, reg (B, EF)	The general conclusion that can be drawn is that the World Bank projections are more accurate than the WEO projections.
Beach et al (1999)	GDP, GDP deflator, CPI, current account	14 countries, 5 country groups	'71-'98	IMF	T, T+1	unclear	reg (B, EF, ED), U, RMSFE, DA	Unbiased and efficient forecasts for developed countries in terms of real GDP growth, inflation, and balance of payments on the current account. For Africa, random walk of real GDP and CPI was more accurate than the WEO forecasts. Overestimates output growth for developing countries.
Keereman (1999)	GDP, inflation, fiscal balance	6 countries	'71-'94	COM,IMF, OECD	T, T+1	first (EC)	MAE, MAE, U, DA	Current year forecast errors for GDP and inflation are similar for the Commission and the IMF.
Batchelor (2000) / Batchelor (2001)	GDP, consumer expenditure, business investment, industrial production, CPI, unemployment	7 countries	'90-'99	IMF, OECD, CONS	T, T+1	settled (CONS)	ME, MAE, RMSFE, ENCOMP	IMF and OECD have in the 1990s been less accurate and less informative than the contemporaneous Consensus Economics forecasts.
Kreinin (2000)	GDP, GDP deflator, unemployment, trade balance	7 countries	'71-'94	IMF, OECD	T, T+1	first (OECD)	reg (B, EF)	IMF real GDP/inflation projections appear superior to the naïve model. Where the OECD and IMF projections do poorly is in forecasting the turning points.
Pons (2000)	GDP	7 countries	'71-'95	IMF, OECD	T, T+1	unclear	reg (B, EF), MAE, RMSFE, DA, U	Summary statistics support the propositions that OECD forecasts are superior to IMF forecasts.
Loungani (2001)	GDP	63 countries	'89-'98	IMF, OECD, WB, CONS	T, T+1	first (WEO)	COR	Correlation between WEO and Consensus forecasts and errors is greater than 0.9.
Artis and Marcellino (2001)	fiscal balance	7 countries	'76-'95	IMF, OECD, COM	T, T+1	first (WEO)	MAE, RMSFE, DM, reg (B, EF), AL, IC, ENCOMP	IMF fiscal forecasts are weakly efficient for the G7 countries with the exception of Japan, but they seem slightly upward biased for Italy, Japan, and the UK, and downward biased for Canada.
Blix et al (2001)	GDP, CPI	6 countries	'91-'00	CONS, OECD, IMF	T, T+1	latest (OECD)	RMSFE, MPE	The IMF and OECD are amongst the worst forecasters for Sweden.
Pons (2001)	GDP deflator	7 countries	'71-'96	IMF	T, T+1	unclear	DA, RAT	The results indicated that the current year forecasts of price increases formulated by the IMF are useful, as on most occasions they correctly indicate the direction of the future evolution of inflation.
Winter WEO (2001)	GDP	7 countries, 7 country groups	'90-'00	IMF, CONS	T, T+1	unclear	ME, SD	Mean and standard deviation of the errors in the WEO for G7 countries are similar to those of private forecasters. Excluding the transition economies, the IMF's global growth forecasts have been relatively unbiased over the past decade.
Golosov and King (2002)	revenues	126 IMF programs (ESAF) / 45 countries	'93-'99	IMF	T+1	first (WEO)	MPE, MAPE, U, DM, reg (B, EF, ED), DA	Evidence of significant serial correlation of the forecast errors.
Hawkins (2002)	GDP, CPI	40 countries	'96-'01	IMF, CONS	unclear	latest (WEO)	MAPE, MPE	Forecasting performance by the IMF is generally comparable to that of the private sector forecasters. IMF forecasts are much less accurate for emerging economies than advanced economies. Forecasts generally do quite a bit better than the naive rule. There is a consistent pattern of over-optimism about economic growth in the emerging economies.
Juhn and Loungani (2002)	GDP	63 countries	'89-'99	IMF, CONS	T, T+1	settled (WEO)	MAE, RMSFE, U, ENCOMP, DA	Evidence tends to favor the private sector forecasts as being a little more accurate than, and encompassing, the WEO forecasts.

[1]See Notes at end of Table.
[2]See Notes at end of Table.

Study	Variables	Countries	Years	Forecasters[1]	Horizon	Actuals	Tests and Stats[2]	Summary of Main Findings (related to IMF forecasts)
Musso and Phillips (2002)	GDP, CPI, current account, capital flows, reserves	69 IMF Programs (SBA, EFF) / 47 countries	'93-'97	IMF	T	unclear	reg (B, EF, ED), DA, U, DM	Three types of projections studies: output growth, and the current and capital accounts of the balance of payments. No statistically significant evidence of bias was found in the main sample considered.
Ashiya (2003)	GDP	7 countries	'85-'01	IMF	T+1, T	first (WEO)	rev, DA	The sign of the difference between the Fall Year ahead and Current Year forecast is useful to predict the acceleration/deceleration of the growth rate of the next year.
GAO (2003)	GDP, CPI, current account	94 countries	'90-'01	IMF	T, T+1	first settled (WEO)	reg (B, EF), U	Forecasts for most developed countries are superior to forecast of emerging market countries when compared to the naïve model forecasts. Forecast for GDP and inflation in program countries are biased, whereas the forecasts for the non-program countries were not. In addition, in all cases the program forecasts were substantially more accurate than the naive model.
Atoian et al (2004)	fiscal balance, current account	175 IMF Programs	'93-'01	IMF	T, T+1, T+2, T+3	latest (WEO)	COR, reg (B, EF, ED), rev, DECOMP, U	While projections outperform a random walk they are not much better. The IMF staff learns from past projection errors--and from new information.
Batista and Zalduendo (2004)	GDP	109 countries	'96-'00	IMF, Model	T+5	unclear	reg (B), RMSFE, U, COR	Model forecasts on have on average a 20 percent lower RMSFE than the projections prepared by IMF staff.
Boettcher (2004)	GDP	7 countries	86-'03	IMF	T, T+1	latest (WEO)	B, rev, RMSFE, reg	The errors in the years 1991-1993 show a subtle overestimation of GDP growth for all G7 countries. Nevertheless the record is mixed. While in the years 1986-1989 GDP growth was underestimated for almost every country, the years 2000-2003 show a large overestimation.
Doepke and Fritsche (2004)	GDP, CPI	Germany	'70-'03	CEA, DIW, IfW, HWWA, ifo, IWH, RWI, COM,IW, IMF, WSI, JWB, OECD, CONS	T, T+1	first (?)	reg (B, EF) RAT, ME, MAE, RMSFE, DA, ENCOMP, DM	For growth forecasts, the IMF autumn forecast seems to have a significantly higher MSFE as compared to most forecasts under investigation due to the fact that it is the earliest annual forecast.
IMF (2004)	GDP, current account, fiscal balance	136 IMF Programs (SBA, EFF, PRGF)	'95-'00	IMF	T-1, T, T+1	unclear	ME, RMSFE, COR, DECOMP	Near-term projections in Fund-supported programs are relatively good... growth is over-predicted by a statistically insignificant 0.4 percentage points. Projection errors for inflation are always under predicting but only statistically significant for year-ahead projections.
Tong (2004)	GDP	16 countries	'96-'04	IMF, EIU	T	unclear	reg (ed), ENCOMP	SDDS subscription improves both the forecast accuracy of the WEO and the EIU reports.
Baqir et al (2005)	GDP, CPI, current account, fiscal balance, primary balance, NDA, NFA, revenues, velocity, expenditures, broad money	94 countries	'89-'02	IMF	T, T+1, T+2, T+3	latest (Article IV)	MAPE, reg (B, EF, ED)	Growth projections are more optimistic in SBAs than in PRGF programs. Projections in high-profile SBAs were more realistic than in other SBAs and PRGFs. Inflation objectives are more optimistic in the SBAs than in the PRGFs.
Elliott et al (2005)	fiscal balance	7 countries	'76-'00	IMF, OECD	T, T+1	unclear	RAT, AL	Find systematic over predictions of government budget deficits. When we allow for asymmetric loss we can no longer reject forecast rationality.
Glueck and Schleicher (2005)	GDP	7 countries	'86-'03	IMF, OECD	T, T+1	latest (OECD)	ME, COR	Strong evidence of common behavior in forecast errors between the U.S. and Canada on the one hand and, between German, France and Italy on the other. Forecast errors have become more synchronized over time.
IMF (2005)	external debt, fiscal balance, revenues, expenditures	136 countries	'90-'04	IMF	T+1, T+3, T+5	unclear	reg (B, ED), DECOMP	While there is a statistically significant positive bias in external debt projections (the actual debt ratio is higher than projected), the bias is small, and seems to arise mainly from errors in the projection of the U.S. dollar value of GDP rather than the debt levels per se. There is no evidence of any greater bias in projections undertaken in the context of Fund-supported programs relative to surveillance cases.
Ashiya (2006)	GDP, GDP deflator	7 countries	'72-'03	IMF, OECD	T, T+1	unclear	rev, reg (B, EF), RAT	Overall, IMF forecasts for Germany, the UK and Italy fare well in the rationality test, while forecasts for Japan, Canada, and France do not.
IEO (2006)	GDP, inflation	6 country groups	'91-'03; '94-'03	IMF, CONS, WB, AsDB, AfDB, ECLAC, OECD	T, T+1	latest (WEO)	reg (B, EF) MAE, RMSFE	A comparison of MAEs and RMSEs between WEO and Consensus forecasts indicate that the two sets of forecasts were very close. A formal test indicates that WEO forecasts were not statistically different from Consensus mean forecasts for almost all countries in the sample. (p40). We may say that, if the IMF forecasts were optimistic for Africa and Latin America, the World Bank, AfDB and ECLAC forecasts were even more so.
Takagi and Kucur (2006)	GDP, inflation	109 countries	'94-'03	IMF	T, T+1	latest (WEO)	reg (B, EF, ED)	The magnitude of the forecast errors generally declined over time. Growth forecasts for industrial countries may have been pessimistic, while forecasts for Africa were clearly optimistic. Growth forecasts were significantly optimistic in program countries, particularly those with PRGF programs. Optimism remains for inflation, but the significant bias for inflation cannot be attributed to the presence of an IMF program.
Timmermann (2006) / Spring WEO (2006) / Timmermann (2007)	GDP, inflation, current account, imports, exports	178 countries	'90-'03 '71-'03	IMF, CONS	T, T+1	first (WEO)	reg (B, EF), DA, rev, RMSFE, ENCOMP	Forecasts for real GDP growth display a tendency for systematic over prediction. Biases are persistent over time. Accuracy problems appear related to the standing WEO assumption that the output gap is eliminated after five years. Bias toward under prediction of inflation significant for African, European and Western Hemisphere countries. Forecast performance of the WEO is similar to that of the Consensus forecasts. Timing of the comparison with the Consensus forecast matters.

[1]See Notes at end of Table.
[2]See Notes at end of Table.

Study	Variables	Countries	Years	Forecasters[1]	Horizon	Actuals	Tests and Stats[2]	Summary of Main Findings (related to IMF forecasts)
Aldenhoff (2007)	GDP, GDP deflator, private consumption deflator, CPI, unemployment	6 country groups	'71-'04	IMF, NIESR, JWB, CEA, OECD, CONS	T, T+1	first (WEO)	ME, reg (B, ED), DM	IMF and OECD forecasts overestimated growth to a significant extent. With regard to Africa, the Western Hemisphere, Central and Eastern Europe and--to a lesser extent--also to the Middle East, the IMF is clearly too optimistic, particularly for year ahead projections.
Melander et al (2007)	GDP	15 countries, 2 country groups	'98-'05	IMF, COM,OECD, CONS	T, T+1	first (EC)	ME, MAE, RMSFE, U, DA, reg (B, EF), ENCOMP	The forecast errors generally seem larger for the forecasts prepared by the IMF, whether measured by ME, MAE, or RMSE, especially for the year-ahead outlook with the exception of Ireland. Differences could be explained by timing.
Ragacs and Schneider (2007)	GDP, Inflation, unemployment, fiscal balance	Austria	'98-'06	OeNB, WIFO, GI, OECD, IMF, COM,Model	T, T+1	first (WIFO)	RMSFE, MAE, ME, U, DA, DM	Credit for the most accurate forecasts is due to the national institutions; none of the international institutions reach or significantly surpass their performance. International institutions do not outperform the national forecasters for any of the four variables, not even once.
Dreher (2008)	GDP, CPI	157 countries	'99-'05	IMF	T, T+1	unclear	ME, reg (B, EF, ED)	Countries voting in line with the US in the UN receive better inflation and--depending on the sample of countries included in the analysis--growth forecasts. Inflation forecasts are systematically biased downwards for countries more heavily indebted to the IMF.
Oesterholm and Zettelmeyer (2008)	GDP	6 countries	'02-'07	IMF, Model	T, T+1	first	RMSFE	For the current year the model does slightly better than WEO, for the year-ahead the model also performs slightly better. The model is more sensitive than the WEO in picking up turning points but exaggerates the cyclical change.
Takagi and Kucur (2008)	GDP, inflation	33 countries	'91-'03	IMF, OECD, CONS	T, T+1	latest (WEO)	ME, RMSFE, herding	The current-year forecasts had an overall pessimistic tendency--indicated by negative average errors for growth and positive average errors for inflation. OECD's current-year forecasts for growth outperformed the others in 19 out of the 23 countries. For inflation, the consensus current-year forecasts outperformed the other forecasts in 14 countries. IMF forecasts only outperformed the others in Netherlands and Portugal. The IMF was the worst performer in 10 of the 23 countries for which all three provided forecasts.
Krkoska and Teksoz (2009)	GDP, inflation	25 countries	'94-'07	IMF, UN, EBRD, COM,OECD, EIU, CS, DB, GI, IWH, JPM, KD, WIIW	T, T+1	first	reg (B, EF), MAE	The IMF statistically significant negative growth bias although it was less than the UN and EBRD (p380), just one out of 13 forecasters (Global Insight) provides both unbiased and efficient forecasts of both inflation and growth.
Vaubel (2009)	GDP	7 countries	'73-'85; '90-'04	IMF, OECD, CONS, NIESR, CEA	T+1 (fall only)	unclear	MAPE	IMF forecasts have the largest errors for a all countries except one. The best forecasters are private institution. Among public forecasters, the local institutes are best. Among the international public institutions compared, the OECD is much better forecaster of economic developments in the industrial countries than the organization of almost universal membership.
Atoian and Conway (2010) / Atoian and Conway (2011)	fiscal balance, current account, GDP per capita	291 IMF Programs	'93-'09	IMF	T, T+1, T+2, T+3	latest (WEO)	DECOMP, ME	IMF staff was apparently working with quite different information about the initial conditions of the program countries than is currently accepted as historical. Policy projection error played only a small role in forecasting error.
Marinheiro (2010)	GDP, fiscal Balance	15 countries	'98-'07	VN, COM,IMF	T, T+1, T+2, T+3	unclear	MAE, ME, RMSFE, ENCOMP, reg(B, EF), DM	IMF forecasts for the current year tend to be pessimistic. Forecasts produced by the two international organizations for thee year-ahead horizon perform better in 2/3 of the cases. A tie is dominant between the IMF and national governments and between the IMF and EC forecasts.
Mrkaic (2010)	GDP, Inflation	180 countries	'99-'07	IMF	T, T+1	first (WEO)	VAR/COVAR, ME, SD, COR, ED	The findings in the paper broadly support the claim that increased SDDS subscription or graduation from GDDS to SDDS will be followed by improvements in the statistical quality of the WEO forecasts.
Novotny and Rakova (2010)	GDP, Inflation	4 Countries	'94-'09	CONS, OECD, IMF	T, T+1	first (CONS)	ME, MAE, MAPE, MSFE, RMSFE, DM, ENCOMP	We are not able to make any strong conclusion about the differences between the Consensus forecasts on the one hand and the IMF and OECD forecasts on the other since they are not statistically significant.
Abreu (2011)	GDP, CPI	9 countries	'91-'09	IMF, COM,OECD, EIU, CONS	T, T+1	first	ME, SD, RMSFE, U, ENCOMP, reg (B, EF), DA	Year-ahead forecast errors are predominantly below zero (overestimation) for most countries and are especially pronounced at the beginning and end of the sample period when most countries were experiencing economic recessions. For inflation projection errors are in general similar for IMF, Consensus and EIU. All forecasters have lower MSE's that the naive forecasts. The quantitative accuracy of the IMF's inflation forecasts is, by and large, similar to that of Consensus or EIU.
Borin et al (2011)	GDP	1 country group	'99-'09	IMF, Model, CONS	T, T+1	unclear	RMSFE, ENCOMP	In order to predict the current year, the World Bridge Model forecasts are generally better than previous WEO ones.
Jagric and Beko (2011)	GDP, inflation	Slovenia	'97-'09	IMF, IMAD, BS, SKEP, COM,WIIW	T, T+1	unclear	ME, MAE, RMSFE, U, DA, reg (B, EF)	Results of the analysis do not provide us with an "absolute winner".
Pisan-Ferry et al (2011)	fiscal balance, GDP	5 countries	'05-'10	IMF, COM	T+1	first (EC)	RMSFE	Overall, for the fiscal deficit and the GDP growth rate, the EC seems to forecast better. In six out of ten cases, the EC outperforms the IMF. And when the IMF performs better, the two forecasts are very close. Only one case was different which was the EC made a huge forecasting error of the Greek deficit in 2009. The GDP forecasts are very close.
Zeng (2011)	GDP	Korea	'00-'08	IMF, Model	Q: T-T+6	latest (CEIC)	RMSFE, U	For both the standard and extended evaluation periods, the benchmark random walk model outperforms WEO at most forecast horizons. When aggregated into annual forecasts, the WEO forecast framework turns out to be the second best forecast model, only next to the AR1 model.

[1]See Notes at end of Table.

Study	Variables	Countries	Years	Forecasters[1]	Horizon	Actuals	Tests and Stats[2]	Summary of Main Findings (related to IMF forecasts)
Cabanillas and Terzi (2012)	GDP, inflation	29 countries	'90-'11	COM,IMF, OECD, CONS	T, T+1	first	ME, MAE, RMSFE	Results are mixed for the current-year forecast. While for some Member States, the Commission clearly outperforms the IMF (i.e. Denmark and Greece), for others (i.e. Ireland and Luxembourg), it is the IMF that outperforms the Commission. For the EU and the euro-area aggregates, the forecast accuracy is similar.
ECB (2012)	inflation	Euro Area	'02-'11	COM,IMF, OECD, CONS, SPF, ECB	T, T+1	unclear	ME	The means of the inflation expectations collected by Consensus Economics, and in particular those collected through the SPF, appear to have performed somewhat better in terms of bias than those of International organizations such as the IMF, OECD and European Commission.
IMAD (2012)	GDP, inflation	Slovenia	'02-'11	IMF, IMAD, BS, SKEP, COM,WIIW	T, T+1	unclear	MAE	For GDP the most optimistic forecasts were the autumn year-ahead forecast by IMAD and the IMF. For inflation the most accurate forecasts were made by IMAD and the IMF in the spring and the IMF and OECD in the autumn.
IMF (2012)	current account, fiscal balance, GDP, inflation, reserves	148 IMF Programs (GRA, PGRT)	'02-'11	IMF	T, T+1, T+2	unclear	MAE, ME	No evidence of optimistic bias in fund projections. In GRA programs there was some evidence for pessimistic projections for reserves and perhaps the current account balance. In PRGT programs, there was some weak evidence both for optimistic projections for growth and for pessimistic projections for the current account balance and reserves.
Julio and Esperanca (2012)	GDP, GDP components	7 countries	'93-'10	IMF, OECD	T, T+1	latest	MTWAE, MTWSE, U, RMSFE, reg (B, EF) DECOMP, ME, MAE	OECD's forecasts for GDP components are more accurate than IMF's forecasts. Forecasts are in general inefficient, both for GDP and its components, but inefficiency is more acute in government consumption predictions.
Sahin (2012)	GDP, inflation	21 countries	'03-'10	IMF, BMI	T+1	n/a	reg (ED)	The IMF forecasts for economic growth are statistically significantly more optimistic than private forecasts except for the September current year forecasts. The IMF's inflation forecasts are also more optimistic than the private forecasts. As forecasts become more biased, investors' reliance on the IMF forecasts decreases, results do not lend support to the argument that politically motivated bias provide extra information to the sovereign bond investors.
Tsuchiya (2012)	GDP, GDP deflator, CPI	Japan	'79-'11	IMF, JG	T, T+1	first (WEO)	RAT	Current year forecasts for all variables are useful. However, year-ahead forecasts for GDP gives false signals.
Blanchard and Leigh(2013) / WEO (2012)	GDP	26 countries	'10-'11	IMF, COM, OECD, EIU	T+1	latest (WEO)	reg (ED)	In advanced economies, stronger planned fiscal consolidation has been associated with lower growth than expected. The baseline result of a negative relation between growth forecast errors and planned fiscal consolidation holds for all forecasters considered buy is strongest for IMF forecasts and for EC forecasts.
Faust (2013)	GDP, inflation	169 countries	'90-'09	IMF, CONS	T, T+1	settled (WEO)	ENCOMP, RMSFE, reg (B, EF)	
Drechsel et al (2013)	GDP	3 regions	'90-'11	IMF, Model	M: T-T+24	first (WEO)	MAFE, U	Comparing the forecast error based on the single indicator forecasts with the forecasts of the IMF, it turns out that with the help of some indicators of the forecast quality can be significantly improved in some forecast rounds.
Frenkel et al (2013)	GDP, inflation, current account, fiscal balance	7 countries	'89-'10	IMF, CONS, OECD	T, T+1	latest (national)	herding	Results show that private forecasters intentionally place their forecasts away from the forecasts published by the IMF and the OECD.
Stekler and Zhang (2013)	GDP, inflation	China	'99-'10	IMF, OECD, CONS	T, T+1	first (WEO)	U, MAE, MSFE, ENCOMP	OECD GDP forecasts encompass WEO forecasts while IMF inflation forecasts encompass OECD. Private sector forecasts of GDP encompass both agencies.

[1]Forecasters: African Development Bank (AfDB), Asian Development Bank (AsDB), Austrian Institute for Economic Research (WIFO), Austrian National Bank (OeNB), Bank of Slovenia (BS), Business Monitor International (BMI), Slovenia Chamber of Commerce and Industry (SKEP), Consensus Economics (CONS), Credit Suisse (CS), Dun & Bradstreet (DB), Economic Commission for Latin America and the Caribbean (ECLAC), Economist Intelligence Unit (EIU), European Bank for Reconstruction and Development (EBRD), European Central Bank (ECB), European Commission (COM), German Council of Economic Advisors (CEA), German Government (JWB), German Institute for Economic Research (DIW), Hamburg Institute for Economic Research (HWWA), IHS Global Insight (GI), IMF World Economic Outlook and Program Forecasts (IMF), Institute for Economic Research Halle (IWH), Institute for the German Economy (IW), Japanese government (JG), JP Morgan (JPM), Kiel Institute for the World Economy (IfW), Kopint-Datorg (KD), IMF Monitoring of Fund Arrangements (MONA), Munich Institute for Economic Research (ifo), UK National Institute of Economic and Social Research (NIESR), OECD Economic Outlook (OECD), Rhine-Westphalia Institute for Economic Research (RWI), Slovenia Institute for Macroeconomic Analysis and Development (IMAD), Survey of Professional Forecasters (SPF), German Trade Union Institute (WSI), United Nations (UN), Various Models (Model), Various Government (VN), Vienna Institute for International Economic Studies (WIIW), World Bank (WB)

[2]Tests and statistics: Asymmetric loss functions (AL), bias tests (B), correlations (COR), Diebold-Mariano test (DM), directional accuracy tests (DA), efficiency tests (EF), error decomposition (DECOMP), error determinants (ED), forecast encompassing tests (ENCOMP), herding tests, intercept correction (IC), mean absolute error (MAE), mean absolute percent error (MAPE), mean error (ME), mean percent error (MPE), mean square forecast error (MSFE), median absolute error (MEDAE), mean total weighted absolute error (MTWAE), mean total weighted squared error (MTWSE), policy response tests (PR), rationality tests (RAT), regression analysis (Reg), revision analysis (Rev), root mean squared forecast error (RMSFE), standard deviation (SD), Thiel's U Statistic (U), variance and covariance (VAR/COVAR)

[2]See Notes at end of Table.

ANNEX 3. SUMMARY OF THE LITERATURE ON FORECAST COMPARISONS

Comparing Public Forecasters

The first paper to examine the performance of the IMF's forecasts vis-à-vis another forecaster is Artis (1988), who compares IMF forecasts against OECD and national forecasts of GDP and inflation from 1973–85.[1] Using root mean square errors and absolute errors, Artis finds that OECD's growth forecasts are "slightly superior" while the IMF does better on balance of payments and inflation forecasts. The paper also uses forecast-encompassing tests to determine whether forecasts from one agency can help explain the other agency's forecast errors, and finds that there is "no unexploited information" that could be useful to either forecaster. In an attempt to account for potentially important differences between the release dates of the *WEO* and national forecasts, Artis divides the data into two samples and finds that for GDP forecasts the IMF performed better in the 1970s when it had an information advantage and worse in the 1980s when it had an information disadvantage, but does not find any differences in the inflation forecasts.

The next paper to compare the IMF's forecasts with those of another international organization was Verbeek (1998), who as a part of his analysis of the World Bank Unified Survey forecasts compares them against the *WEO*'s forecasts of GDP, inflation, domestic investment, government deficits, trade, and current accounts for 23 developing countries from 1991–97.[2] Though acknowledging that most World Bank economists "prefer to use the IMF data and projections as input in their own economic work," Verbeek finds that in general the *WEO* forecasts have higher root mean square errors and mean absolute errors than the World Bank forecasts for all variables except government deficits. The paper concludes that the World Bank Unified Survey is more accurate than the *WEO* for both the current-year and year-ahead forecasts.

Several studies compare the accuracy of the OECD and *WEO* forecasts for the G7 countries despite slight differences in release dates and the potential effect that this might have on their relative performance. In his examination of GDP, inflation, unemployment, and trade forecasts from 1971–94, Kreinin (2000) finds that the OECD and IMF forecasts are equally bad at forecasting turning points. Comparing GDP forecasts from 1971–95, Pons (2000) finds that OECD forecasts are superior to IMF forecasts. Elliott and others (2005) examine forecasts of government deficits using asymmetric loss functions and find greater bias in the IMF forecasts. Glück and Schleicher (2005) find that the OECD performs slightly better than

[1] The OECD forecasts come from its semi-annual publication, *Economic Outlook,* which is typically released in June and December.

[2] It is unclear how exactly this comparison was done since only the aggregate developing country group forecasts were publicly available from the *WEO* during this time, and these did not necessarily have the same country composition as those chosen for Verbeek's study. Also, Verbeek uses World Bank data as the actual data.

the IMF in their analysis of GDP forecasts and revisions from 1996–2003. Similarly, when Julio and Esperanca (2012) examine GDP component forecasts from 1993–2010, they find that the OECD's forecasts are slightly more accurate.

Another forecaster that the IMF is frequently compared against is the European Commission (EC). Artis and Marcellino (2001), in the first paper to compare the two forecasters, analyze forecasts of budget deficits for G7 countries from 1976–95. Focusing on the root mean square forecast errors and absolute forecast errors, their paper finds that the IMF performs better for France and Germany while the EC does better for Italy and the U.K. However, it finds that in general the EC performs better than the IMF in forecast-encompassing tests and argues that this result might be explained by the differences in timing between the forecasts or by differences in how forecasters interpret current policies.

The next paper to compare the IMF and EC forecast performance is Melander and others (2007) as a part of their comprehensive analysis of EC forecasts. Using standard summary statistics to examine forecasts of GDP for 15 European countries from 1998–2005, the authors find that except for Ireland and Greece the IMF has larger forecast errors, especially for Germany.[3] From forecast-encompassing tests, they find that while the EC encompasses the IMF forecasts for some countries, the EC forecasts for Austria, Belgium, Greece, and France could be improved by information from the IMF forecasts. Overall, they argue that the EC has an informational advantage given the later release date of its forecasts and that the main difference between the forecasts is the weak performance of the IMF on its year-ahead forecasts for Germany.

Pisani-Ferry and others (2011) update Melander and others (2007)'s analysis of GDP and fiscal deficit forecasts for Greece, Ireland, Portugal, and Spain through 2010. By examining the root mean square errors of the forecasts, they find that although the IMF forecasts have improved relative to the EC's, the EC performs better than the IMF in six out of ten cases. Furthermore, they argue that even though the EC made a "huge forecasting error" for the Greek deficit in 2009, the IMF's underperformance stems mainly from its fiscal projections. Cabanillas and Terzi (2012) provide an update through 2011. For current-year forecasts for EU member states they find mixed results, but for year-ahead forecasts they find that the EC performs better, and note that the EC has an informational advantage.

Comparing Private Forecasters

The first release of Consensus Economics forecasts in 1989 sparked a number of studies comparing the performance of these forecasts against the *WEO* and other official forecasts. The first study to compare *WEO* and Consensus forecasts is Artis (1997), which looks at forecasts of GDP and inflation for G7 countries from 1990–94. Given the short sample period, the paper plots the errors for both forecasts against one another and finds that both

[3] Use European Commission data as actual data.

forecasters make similar errors at the same times for the same countries. Loungani (2001) updates this analysis by looking at forecasts of GDP for 63 developing and advanced economies from 1989–98. He finds that the correlation between the forecast errors is greater than 0.9 and concludes that differences between them would be statistically insignificant.

Batchelor (2001) finds somewhat different results. His paper compares forecasts of GDP, inflation, consumer expenditure, and unemployment for a narrower set of G7 countries from the *WEO* and Consensus Economics from 1990–99.[4] He finds that the Consensus forecasts are consistently less biased than *WEO* forecasts for all variables except current-year forecasts of inflation and that Consensus forecasts are able to significantly encompass the information in the IMF forecasts.

Additional studies focusing on comparisons between *WEO* and Consensus Economics forecasts have continued to find mixed results on relative forecast performance. A box in the Winter 2001 edition of the *World Economic Outlook* acknowledges that from 1990–2000, Consensus' average error is slightly smaller for the year-ahead forecasts, but that its standard deviation is also slightly higher, which makes it difficult to say which forecast is better. Juhn and Loungani (2002) examine forecasts for 63 countries from 1989–99 and find that Consensus performs slightly better than, and encompasses, the IMF's forecasts. Hawkin's (2002) analysis of GDP and inflation forecast for 40 countries from 1996–2001 finds that the performance of the IMF forecasts is largely comparable to that of the Consensus forecasts. Similarly, Timmermann (2007) compares GDP, inflation, and current account forecasts for 23 countries from 1990–2003 and finds that neither forecaster completely outperforms the other, although it does matter which vintage of Consensus forecasts one considers.[5]

Comparing Groups of Forecasters

A wide variety of studies compare the *WEO* forecasts with those from other forecasters, yielding often contradictory results. Several of these studies find serious problems in the IMF's forecasts. For example, Blix and others (2001) compare IMF forecasts of GDP and inflation for six advanced economies from 1991–2000 with those from the OECD and a variety of private forecasters and find the IMF is among the worst forecasters for Sweden. Takagi and Kucur (2008) compare the *WEO*, OECD, and Consensus Economics forecasts of GDP and inflation for 23 countries from 1991–2003 and find that the IMF is the worst forecaster for almost half the countries in the sample.

Other studies are less critical but find specific evidence that in year-ahead forecasts the IMF is more biased than other forecasters. In their analysis of forecasts of Germany's GDP and

[4] Analysis also includes OECD forecasts which performed similarly to the IMF.

[5] Since Consensus Economics forecasts are produced monthly it is not clear which month should be used to compare against the *WEO* forecasts. Timmermann (2006, 2007) finds that the selection of different months does affect their relative forecast performance.

inflation, Doepke and Fritsche (2004) find that the IMF's year-ahead forecasts are worse than other agencies', mainly because they are produced first. Aldenhoff (2007) confirms the poor performance of the IMF's year-ahead forecasts for different groups of advanced and emerging markets in his analysis of various public and private GDP, inflation and unemployment forecasts from 1971–2003. Vaubel (2009) confirms this finding for GDP forecasts in G7 countries from various forecasters.

However, not all studies find that IMF performs any worse than other forecasters. As part of its evaluation of the IMF's multilateral surveillance, IEO (2006) compares *WEO*, Consensus, World Bank, OECD, and regional development bank forecasts of GDP and inflation for different regions from 1991–2003 and finds that *WEO* forecasts do just as well as or better than other forecasters. Krkoska and Teksoz (2009) compare forecasts of GDP and inflation for 25 countries from 13 different international, national, and private sources for the period of 1994–2007 and find that while the IMF's GDP forecasts are biased, they are less so than those produced by other international organizations and that no single forecaster outperforms the others in all areas. Abreu (2011) compares the IMF, EC, OECD, Consensus Economics, and Economist Intelligence Unit (EIU) forecasts of GDP and inflation for nine European countries from 1991–2009 and finds that in general the forecasts of international organizations do not differ significantly from those of the private sector. Jagric and Beko (2011) examine forecasts of inflation and GDP for Slovenia from 1997–2009 from various sources and find that while the IMF makes the best year-ahead forecasts, especially when 2009 is included, overall none of the forecasters outperforms any of the others.

ANNEX 4. INTERPRETATION OF THE REGRESSION RESULTS IN TABLE 4[1]

Suppose that growth in country i (y^i) can be expressed as linear function of growth in China, Germany, and the United States as well as of a set of other variables denoted X as in equation (1) below. The size of the spillovers from the three large economies is given by the coefficients β. We assume that the IMF forecaster knows the relationship except possibly for the size of the spillover coefficients. Hence the forecast of the IMF will be based on equation (2).

$$y_t^i = \beta_{i,CN}^{TRUE} y_t^{CN} + \beta_{i,GE}^{TRUE} y_t^{GE} + \beta_{i.US}^{TRUE} y_t^{US} + \gamma_i X_t^i + \varepsilon_t^i \quad (1)$$

$$y_t^{i,IMF} = \beta_{i,CN}^{IMF} y_t^{CN} + \beta_{i,GE}^{IMF} y_t^{GE} + \beta_{i.US}^{IMF} y_t^{US} + \gamma_i X_t^i + \varepsilon_t^i \quad (2)$$

Denoting forecasts by the IMF with a circumflex ($\hat{}$) it is easy to show that the forecast error of growth in country i will be a function of the forecasts of growth in China, Germany, and the United States as in equation (3).

$$y_{t+1}^i - \hat{y}_{t+1}^i = (\beta_{CN}^{TRUE} - \beta_{CN}^{IMF})\hat{y}_{t+1}^{CN} + (\beta_{GE}^{TRUE} - \beta_{GE}^{IMF})\hat{y}_{t+1}^{GE} + (\beta_{US}^{TRUE} - \beta_{US}^{IMF})\hat{y}_{t+1}^{US} + u_t^i \quad (3)$$

where

$$u_t^i = \beta_{i,CN}^{TRUE}(y_{t+1}^{CN} - \hat{y}_{t+1}^{CN}) + \beta_{i,GE}^{TRUE}(y_{t+1}^{GE} - \hat{y}_{t+1}^{GE}) + \beta_{i.US}^{TRUE}(y_{t+1}^{US} - \hat{y}_{t+1}^{US}) + \gamma_i\left(X_{t+1}^i - \hat{X}_{t+1}^i\right) + \varepsilon_{t+1}^i \quad (4)$$

Following Timmermann (2006) we estimated a regression equation of the form (5) in which the IMF's forecast error for the growth rate of country i is a linear function of the IMF's forecasts of the growth rate in China, Germany, and the United States and an error term.

$$y_{t+1}^i - \hat{y}_{t+1}^{i,IMF} = a_{CN}\hat{y}_t^{CN,IMF} + a_{GE}\hat{y}_t^{GE,IMF} + a_{US}\hat{y}_t^{US,IMF} + v_t^i \quad (5)$$

Comparing (5) with (3) it is clear that the estimated coefficient on the forecast of China (Germany, United States) will be positive if the true spill-over from China (Germany, United States) to country i is larger than the spill-over assumed by the IMF forecaster. It will be negative if the IMF forecaster overestimates the spillover effect.

Comparing (5) with (3) and (4) it is also clear that the error term in (5) will be uncorrelated with the regressors if the forecast errors of growth in China, Germany, and the United States are uncorrelated with the forecasts themselves.

A full country-by-country analysis of the regression results based on this interpretation of the regression coefficients is beyond the scope of this paper, but a few generalizations can

[1] For Table 4, see page 15.

nevertheless be made. First, the IMF tends to underestimate the spill-over effect from the US more frequently than it tends to overestimate this effect. This is particularly the case in the 2000–11 period. The opposite is the case for China, overestimates of the spill-over from China is more frequent than underestimates, albeit only marginally. Second, the IMF tends to overestimate the spill-overs from China in CIS countries and Mongolia, and tends to underestimate its spill-overs to Africa. Third, the IMF has tended to underestimate the spill-overs from Germany to CIS countries and Mongolia, especially during the 1990-2003 period.

ANNEX 5. COUNTRY COMPARISONS WITH CONSENSUS FORECASTS

Country	Spring Year-Ahead			Fall Year-Ahead			Spring Current-Year			Fall Current-Year		
	Time Period	RMSFE CON/WEO	RMSFE DM (pval)	Time Period	RMSFE CON/WEO	RMSFE DM (pval)	Time Period	RMSFE CON/WEO	RMSFE DM (pval)	Time Period	RMSFE CON/WEO	RMSFE DM (pval)
African Department												
Nigeria	'00-'11	0.90	0.46	'00-'11	1.06	0.20	'99-'11	1.14	0.47	'99-'11	1.09	0.49
South Africa	'95-'11	1.17	0.06 *	'94-'11	1.12	0.00 ***	'94-'11	1.11	0.02 **	'93-'11	1.02	0.83
Asia Pacific Department												
Australia	'91-'11	0.88	0.02 **	'91-'11	1.00	0.94	'91-'11	0.81	0.02 **	'91-'11	0.75	0.24
Bangladesh	'96-'11	1.15	0.23	'96-'11	0.88	0.49	'95-'11	1.16	0.04 **	'95-'11	1.23	0.16
China	'96-'11	1.06	0.49	'96-'11	1.02	0.92	'95-'11	0.97	0.39	'95-'11	0.71	0.10
Hong Kong SAR	'92-'11	1.00	0.92	'92-'11	1.06	0.16	'91-'11	0.96	0.22	'91-'11	1.03	0.60
India	'96-'11	0.91	0.34	'96-'11	0.94	0.28	'95-'11	1.06	0.23	'95-'11	1.13	0.29
Indonesia	'92-'11	1.00	0.79	'92-'11	1.01	0.40	'91-'11	0.85	0.09 *	'91-'11	1.15	0.34
Japan	'91-'11	0.97	0.22	'91-'11	1.00	0.99	'91-'11	0.91	0.07 *	'91-'11	1.15	0.00 ***
Korea	'91-'11	1.06	0.31	'91-'11	0.98	0.43	'91-'11	0.88	0.10 *	'91-'11	0.82	0.25
Malaysia	'92-'11	0.97	0.22	'92-'11	0.99	0.65	'91-'11	0.87	0.01 **	'91-'11	0.63	0.01 **
New Zealand	'91-'11	0.88	0.15	'93-'11	0.81	0.16	'91-'11	0.74	0.09 *	'92-'11	0.73	0.28
Philippines	'96-'11	0.86	0.17	'96-'11	1.05	0.14	'95-'11	1.01	0.65	'95-'11	1.23	0.01 **
Singapore	'92-'11	0.94	0.45	'92-'11	1.01	0.74	'91-'11	0.86	0.06 *	'91-'11	0.87	0.36
Sri Lanka	'96-'11	0.86	0.01 ***	'96-'11	0.93	0.07 *	'95-'11	0.86	0.02 **	'95-'11	0.89	0.25
Taiwan Province of China	'91-'11	0.90	0.27	'91-'11	1.05	0.37	'91-'11	1.01	0.76	'91-'11	1.05	0.61
Thailand	'92-'11	0.94	0.03 **	'92-'11	0.93	0.16	'91-'11	0.93	0.26	'91-'11	0.93	0.34
Vietnam	'96-'11	0.98	0.63	'96-'11	0.93	0.30	'95-'11	1.21	0.23	'95-'11	1.50	0.15
European Department												
Austria	'91-'11	0.99	0.84	'91-'11	1.05	0.62	'91-'11	1.17	0.02 **	'91-'11	0.80	0.31
Belarus	'00-'11	0.97	0.67	'99-'11	1.01	0.82	'99-'11	0.96	0.46	'98-'11	0.94	0.67
Belgium	'91-'11	1.03	0.67	'91-'11	1.10	0.33	'91-'11	1.01	0.83	'91-'11	1.02	0.88
Bulgaria	'96-'11	1.03	0.01 ***	'96-'11	1.00	0.98	'95-'11	1.05	0.01 ***	'95-'11	0.91	0.31
Croatia	'00-'11	1.09	0.08 *	'99-'11	1.09	0.01 ***	'99-'11	1.30	0.03 **	'98-'11	0.97	0.85
Cyprus	'06-'11	1.06	0.03 **	'05-'11	1.12	0.03 **	'05-'11	1.15	0.17	'04-'11	1.22	0.24
Czech Republic	'96-'11	0.99	0.80	'96-'11	1.04	0.40	'95-'11	0.99	0.93	'95-'11	0.73	0.08 *
Denmark	'91-'11	1.08	0.18	'91-'11	1.01	0.86	'91-'11	1.11	0.32	'91-'11	0.90	0.18
Estonia	'00-'11	1.03	0.17	'99-'11	1.03	0.41	'99-'11	1.08	0.01 **	'98-'11	1.02	0.81
Euro Area	'05-'11	1.02	0.41	'05-'11	1.10	0.04 **	'04-'11	1.10	0.16	'04-'11	0.96	0.62
Finland	'91-'11	1.02	0.24	'91-'11	1.06	0.01 ***	'91-'11	1.13	0.19	'91-'11	1.33	0.00 ***
France	'91-'11	1.02	0.43	'91-'11	1.02	0.74	'91-'11	0.95	0.25	'91-'11	1.08	0.10
Germany	'91-'11	0.94	0.02 **	'91-'11	0.98	0.63	'91-'11	0.97	0.67	'91-'11	1.07	0.43
Greece	'95-'11	1.08	0.34	'94-'11	1.12	0.14	'94-'11	0.88	0.37	'93-'11	1.09	0.31
Hungary	'92-'11	0.95	0.57	'92-'11	0.90	0.41	'91-'11	0.47	0.23	'91-'11	0.96	0.34
Ireland	'91-'11	1.02	0.40	'91-'11	1.18	0.01 ***	'91-'11	1.21	0.00 ***	'91-'11	1.14	0.14
Israel	'96-'11	0.95	0.30	'96-'11	0.95	0.57	'95-'11	0.93	0.44	'95-'11	1.14	0.17
Italy	'91-'11	1.02	0.61	'91-'11	1.08	0.21	'91-'11	1.01	0.90	'91-'11	1.02	0.81
Latvia	'00-'11	1.20	0.09 *	'99-'11	1.14	0.13	'99-'11	1.10	0.14	'98-'11	1.09	0.50
Lithuania	'00-'11	1.01	0.04 **	'00-'11	1.13	0.09 *	'99-'11	1.14	0.00 ***	'99-'11	0.94	0.21
Netherlands	'91-'11	1.03	0.12	'91-'11	1.04	0.07 *	'91-'11	1.00	0.93	'91-'11	1.04	0.54
Norway	'91-'11	0.95	0.50	'91-'11	1.01	0.95	'91-'11	0.96	0.74	'91-'11	1.15	0.28
Poland	'92-'11	0.90	0.05 **	'92-'11	1.02	0.87	'91-'11	0.72	0.24	'91-'11	1.28	0.34
Portugal	'91-'11	0.99	0.82	'91-'11	1.03	0.80	'91-'11	1.02	0.84	'91-'11	1.04	0.86
Romania	'96-'11	1.03	0.10	'96-'11	1.02	0.06 *	'95-'11	1.17	0.05 **	'95-'11	0.91	0.11
Russia	'96-'11	0.90	0.25	'96-'11	0.92	0.47	'95-'11	0.86	0.33	'95-'11	1.08	0.11
Slovak Republic	'96-'11	1.01	0.90	'96-'11	1.04	0.49	'95-'11	1.19	0.11	'95-'11	1.11	0.29
Slovenia	'96-'11	1.08	0.21	'96-'11	1.05	0.09 *	'95-'11	1.07	0.26	'95-'11	0.88	0.32
Spain	'91-'11	1.02	0.34	'91-'11	1.15	0.03 **	'91-'11	1.08	0.21	'91-'11	0.82	0.15
Sweden	'91-'11	1.05	0.08 *	'91-'11	1.02	0.27	'91-'11	0.96	0.56	'91-'11	1.04	0.79
Switzerland	'91-'11	0.98	0.52	'91-'11	1.05	0.03 **	'91-'11	1.01	0.84	'91-'11	0.96	0.59
Turkey	'96-'11	1.01	0.75	'96-'11	1.06	0.01 ***	'95-'11	1.02	0.66	'95-'11	0.89	0.37
Ukraine	'96-'11	0.96	0.71	'96-'11	0.93	0.64	'95-'11	0.91	0.00 ***	'95-'11	1.06	0.71
United Kingdom	'91-'11	1.00	0.94	'91-'11	1.08	0.19	'91-'11	1.15	0.08 *	'91-'11	0.97	0.74
Middle East and Central Asia Department												
Azerbaijan	'00-'11	1.24	0.13	'99-'11	1.03	0.71	'99-'11	1.24	0.06 *	'98-'11	1.12	0.54
Egypt	'96-'11	1.06	0.51	'96-'11	1.11	0.03 **	'95-'11	1.16	0.28	'95-'11	1.11	0.68
Kazakhstan	'00-'11	0.96	0.00 ***	'99-'11	1.00	0.90	'99-'11	0.99	0.57	'98-'11	0.92	0.36
Pakistan	'96-'11	0.89	0.16	'96-'11	0.89	0.25	'95-'11	1.35	0.03 **	'95-'11	0.82	0.03 **
Saudi Arabia	'96-'11	1.04	0.64	'96-'11	1.11	0.06 *	'95-'11	0.76	0.00 ***	'95-'11	1.23	0.00 ***
Uzbekistan	'00-'11	0.84	0.04 **	'99-'11	0.95	0.67	'99-'11	0.93	0.47	'98-'11	1.03	0.88
Western Hemisphere Department												
Argentina	'94-'11	0.99	0.56	'95-'11	0.97	0.29	'93-'11	0.82	0.00 ***	'94-'11	0.99	0.85
Bolivia	'94-'11	0.77	0.12	'95-'11	0.78	0.16	'93-'11	0.90	0.65	'94-'11	0.74	0.40
Brazil	'91-'11	0.91	0.14	'95-'11	1.02	0.70	'91-'11	0.97	0.75	'94-'11	0.84	0.01 ***
Canada	'91-'11	0.91	0.29	'91-'11	1.03	0.54	'91-'11	0.93	0.20	'91-'11	0.94	0.33
Chile	'94-'11	0.98	0.46	'95-'11	1.01	0.79	'93-'11	1.03	0.79	'94-'11	0.98	0.80
Colombia	'94-'11	1.02	0.33	'95-'11	1.15	0.04 **	'93-'11	1.06	0.06 *	'94-'11	0.88	0.35
Costa Rica	'94-'11	0.98	0.71	'95-'11	0.98	0.57	'93-'11	0.93	0.00 ***	'94-'11	1.07	0.34
Dominican Republic	'94-'11	1.11	0.00 ***	'95-'11	1.07	0.06 *	'93-'11	1.01	0.87	'94-'11	1.06	0.61
Ecuador	'94-'11	0.93	0.05 *	'95-'11	1.08	0.09 *	'93-'11	1.23	0.09 *	'94-'11	1.42	0.02 **
Mexico	'91-'11	1.00	0.98	'95-'11	1.02	0.66	'91-'11	0.91	0.08 *	'94-'11	1.05	0.54
Panama	'94-'11	0.97	0.65	'95-'11	1.01	0.84	'93-'11	0.96	0.71	'94-'11	0.91	0.58
Paraguay	'94-'11	0.89	0.00 ***	'95-'11	0.99	0.87	'93-'11	1.03	0.70	'94-'11	1.12	0.53
Peru	'94-'11	1.04	0.56	'95-'11	0.95	0.46	'93-'11	1.15	0.41	'94-'11	1.25	0.24
United States	'91-'11	1.06	0.23	'91-'11	1.07	0.50	'91-'11	0.93	0.12	'91-'11	1.10	0.00 ***
Uruguay	'94-'11	0.96	0.11	'95-'11	0.94	0.11	'93-'11	1.10	0.00 ***	'94-'11	1.12	0.01 ***
Venezuela	'94-'11	1.04	0.36	'95-'11	1.01	0.67	'93-'11	1.02	0.85	'94-'11	0.82	0.22

ANNEX 6. RECESSION EPISODES

Australia '91
Austria '93, '09
Belgium '93, '09
Canada '91, '09
Cyprus '09
Czech Republic '09
Denmark '08, '09
Euro Area '09
Finland '91, '92, '93, '09
France '93, '09
Germany '93, '03, '09
Greece '09, '10[§], '11[§]
Hong Kong SAR '98, '09
Iceland '92, '02, '09[§], '10[§]
Ireland '08, '09, '10
Israel '01, '02
Italy '93, '05, '08, '09
Japan '98, '01, '08, '09, '11
Korea '98
Luxembourg '09
Malta '09
Netherlands '03, '09
New Zealand '98, '09
Norway '09
Portugal '93, '03, '08, '09, '11[§]
Singapore '01, '09
Slovak Republic '09
Slovenia '09
Spain '93, '09, '10
Sweden '91, '92, '93, '08, '09
Switzerland '91, '92, '93, '96, '03, '09
Taiwan Province of China '01, '09
United Kingdom '91, '92, '09
United States '91, '09

Algeria '93, '94
Antigua and Barbuda '95, '01, '09, '10[§], '11[§]
Argentina '95[§], '99[§], '00[§], '01[§], '02[§]
Barbados '91, '92[§], '01, '02, '09
Belarus '92, '93, '94, '95[§]
Belize '09
Bosnia and Herzegovina '09[§]
Botswana '09
Brazil '92[§], '98, '03[§], '09
Brunei Darussalam '08, '09
Bulgaria '91[§], '92, '93, **'96, '97[§], '09**
Chile '99, '09
Colombia '99
Costa Rica '96, '09
Croatia '93, **'99[§], '09, '10, '11**
Czech Republic '93, **'98, '99**
Dominican Republic '03
Ecuador '99
El Salvador '09[§]
Equatorial Guinea '91[§], '06, '10
Estonia '92[§], '93, **'99, '08, '09**
Fiji '07, '09
FYR Macedonia '09
Gabon '99, '00, '09[§]
Guinea '09[§]
Hungary '91[§], '92[§], '93[§], '09[§]
Indonesia '98[§]
Iraq '91, '96
Jamaica '96, '97, '98, '08, '09, '10[§]
Kazakhstan '92, '93, '94[§], '95[§], **'98[§]**
Kuwait '91, '99, '01, '02, '09
Latvia '92[§], '93, **'08, '09[§], '10[§]**
Lebanon '99, '00, '06
Libya '93, '94, '95, '02, '09, '11
Lithuania '92, '93, **'99, '09**
Malaysia '98, '09
Malta '01, '03
Mexico '95[§], '01, '09
Montenegro '09
Morocco '92, '95, '97, '99
Namibia '93, '00, '09
Papua New Guinea '95[§], '97[§], '01[§], '02
Paraguay '00, '02, '09
Peru '92
Philippines'98[§]
Poland '91[§]
Qatar '95
Romania '91[§], '92, '93, **'97[§], '98, '99, '09[§], '10[§]**
Russia '92, '93, '94, **'95, '96[§], '98[§], '09**
Saudi Arabia '94, **'95, '99**
Serbia '09[§]
Seychelles '94, '95, '01, '03, '04, '05, '08
Slovak Republic '93
South Africa '91, '92, **'09**
St. Kitts and Nevis '09, '10, '11[§]
Suriname '91, '92, '93, '94
Swaziland '92
Syria '99
Thailand '97, '98[§], '09
The Bahamas '91, '92, '01, '08, '09
Trinidad and Tobago '93, '09, '10, '11
Turkey '94[§], '99, '01[§], '09
Turkmenistan '92, '93, '94, '95, '96, '97
Tunisia '11
Tuvalu '10
Ukraine '92, '93, '94, **'95[§], '96, '97, '98, '99[§], '09[§]**
United Arab Emirates '09
Uruguay '95, '99, '00[§], '01[§], '02[§]
Uzbekistan '92, '93, '94, '95
Venezuela '93, '94, '96[§], '98, '99, '02, '03, '09, '10

Afghanistan '94
Albania '92[§], '97
Angola '93, '98
Armenia '92, '93, **'09[§]**
Azerbaijan '92, '93, '94, '95
Bosnia and Herzegovina '93,
Burundi '93[§], '94[§], '95, '96, '99, '03
Cambodia '09
Cameroon '91, '92, '93, '94[§]
Central African Republic '92, '93, '96, '03
Chad '93, '99[§], '08, '09
Comoros '96, '97, '00
Côte d'Ivoire '91[§], '93, '00[§], '02[§], '03[§], '11[§]
DR Congo '91, '92, '93, '94, '95, '97, '98, '99, '00, '01
Democratic Republic of Timor-Leste '03, '06
Djibouti '92, '93, '94, '95, '96[§]
Dominica '02, '03, '09
Dominican Republic '91[§]
Eritrea '00
Ethiopia '91, '92, '98[§], '03[§]
FYR Macedonia '93, '94, '95[§], '01[§]
Georgia '92, '93, '94, **'09[§]**
Grenada '92, '95, '02, '04, '09[§], '10[§]
Guinea-Bissau '98[§], '02
Guyana '98[§], '03[§], '05[§]
Haiti '92, '93, '94, '01, '02, '04, '10[§]
Honduras '94[§], '99[§], '09
Kenya '00[§]
Kiribati '04, '09
Kyrgyz Republic '92, '93[§], '94[§], '02[§], '05[§], '10[§]
Lesotho '98
Madagascar '91[§], '02[§], '09[§]
Malawi '92[§], '94, '01[§]
Maldives '05, '09
Mali '91[§], '93[§]
Mauritania '09[§]
Moldova '92, '93, '94[§], '95, '96[§], **'98[§], '99[§], '09[§]**
Mongolia '92[§], '93[§], '09[§]
Mozambique '92[§]
Nepal '91, '02
Nicaragua '91[§], '93, **'09[§]**
Niger '92, '09[§]
Philippines '91[§]
Republic of Congo '93, '94[§], '99[§], '07[§]
Rwanda '91[§], '93[§], '94
Samoa '92, '94, '09, '10
Senegal '93
Sierra Leone '91, '92, '95[§], '99
Solomon Islands '97, '99, '00, '01, '02, '09
Sri Lanka '01[§]
St. Lucia '02, '09
St. Vincent and the Grenadines '09, '10
Sudan '11
Tajikistan '92, '93, '94, '95, '96[§]
The Gambia '94, '95, '02[§]
Togo '91[§], '93[§], '98[§], '00
Tonga '96, '97, '98, '07, '09
Vanuatu '99, '01, '02
Yemen '11[§]
Zambia '92, '95, '98[§]
Zimbabwe '92[§], '95[§], '99[§], '00[§], '01, '02, '03, '04, '05, '06, '07, '08

Bold = Consensus forecasts available, [§] IMF Program

ANNEX 7. SAMPLE USED IN STAFF EXPERIENCE ANALYSIS[1]

Afghanistan	Comoros*	Hong Kong SAR	Moldova	Sri Lanka
Albania	Costa Rica	Hungary	Mongolia	St. Kitts & Nevis
Algeria	Côte d'Ivoire	Iceland	Montenegro	St. Lucia
Angola	Croatia	India	Mozambique	St. Vincent & the Grenadines
Antigua & Barbuda	Czech Republic	Indonesia	Myanmar	Sudan*
Argentina*	DR Congo	Ireland	Namibia*	Suriname
Armenia	Timor-Leste	Israel	Nepal	Swaziland*
Australia	Denmark	Italy	Netherlands	Sweden
Austria	Djibouti	Jamaica	New Zealand	Switzerland
Azerbaijan*	Dominica	Japan	Nicaragua	Syria
Bangladesh	Dominican Republic	Jordan	Niger*	Tajikistan
Barbados	Ecuador	Kazakhstan	Norway*	Tanzania
Belarus	Egypt	Kenya	Pakistan*	Thailand
Belgium	El Salvador	Korea	Panama	The Bahamas
Belize	Equatorial Guinea*	Kosovo	Papua New Guinea	The Gambia
Benin	Eritrea	Kuwait	Paraguay	Togo*
Bhutan	Estonia	Kyrgyz Republic	Peru	Trinidad & Tobago
Bolivia	Ethiopia	Lao P.D.R.	Philippines	Turkey
Bosnia & Herzegovina	Fiji*	Latvia	Portugal	Turkmenistan
Botswana*	Finland	Lebanon	Qatar*	Uganda
Brazil	France	Lesotho	Republic of Congo	Ukraine
Brunei Darussalam*	FYR Macedonia	Liberia	Russia	United Arab Emirates
Bulgaria	Gabon	Libya	Rwanda	United Kingdom
Burkina Faso	Georgia	Lithuania	São Tomé & Príncipe*	United States
Burundi	Germany	Luxembourg	Saudi Arabia	Uruguay
Cambodia	Ghana	Madagascar	Senegal	Uzbekistan
Cameroon	Greece	Malawi	Serbia	Vanuatu*
Canada	Grenada	Malaysia	Seychelles	Vietnam
Cape Verde	Guatemala	Maldives	Sierra Leone	Yemen
Central African Republic	Guinea	Mali	Singapore	Zambia
Chad*	Guinea-Bissau*	Malta	Slovak Republic	Zimbabwe
Chile	Guyana	Mauritania	Slovenia	
China	Haiti	Mauritius*	Solomon Islands	
Colombia	Honduras	Mexico	Spain	

[1] Not all countries are included in all samples. Countries are included at least once in at least one sample. * Indicate countries that only occur in the expanded sample.

ANNEX 8. DERIVATIONS OF FORECAST ERROR VARIANCES ACCOUNTED FOR BY GLOBAL AND REGIONAL FACTORS

Consider the following decomposition of year-over-year real GDP growth at time *t* in country *i*, belonging to region *j* ($X_{ij,t}$):[1]

$$X_{ij,t} = A_{ij,0}G_t + A_{ij,1}G_{t-1} + B_{ij,0}R_{j,t} + B_{ij,1}R_{j,t-1} + \Psi_{ij,t} \quad (1)$$

where *G*, the global factor, R_j, the regional factor *j*, and Ψ_{ij}, the idiosyncratic factor for country *ij*, evolve according to:

$$\begin{aligned} G_t &= CG_{t-1} + \mu_t^G \\ R_{j,t} &= D_j R_{j,t-1} + \mu_{j,t}^R \\ \Psi_{ij,t} &= E_{ij}\Psi_{ij,t-1} + \mu_{ij,t}^I \end{aligned} \quad (2)$$

and where

$$\mu_t^G, \qquad \mu_{j,t}^R, \qquad and\ \mu_{ij,t}^I$$

are serially and mutually uncorrelated random variables for all *i* and *j*.

A common measure of the importance of a factor in accounting for growth in a given country is the variance of the forecast error of growth accounted for by this factor. For the model described by equations (1) and (2) this measure can be calculated relatively straightforwardly. Dropping *i, j*, and *t* subscripts for simplicity we obtain the following expression for the forecast variances attributable to the global and regional factors respectively:

$$FEV(G) = \frac{1}{1+\frac{(B_0^2(D^2+1)+B_1^2)\sigma_R^2}{(A_0^2(C^2+1)+A_1^2)\sigma_G^2}+\frac{(E^2+1)\sigma_I^2}{(A_0^2(C^2+1)+A_1^2)\sigma_G^2}} \quad (3)$$

and

$$FEV(R) = \frac{1}{\frac{(A_0^2(C^2+1)+A_1^2)\sigma_G^2}{(B_0^2(D^2+1)+B_1^2)\sigma_R^2}+1+\frac{(E^2+1)\sigma_I^2}{(B_0^2(D^2+1)+B_1^2)\sigma_R^2}} \quad (4)$$

These expressions have intuitive interpretations: A higher value of σ_G^2—the variance of the shocks to the global factor—and larger values of A_0 and A_1—the importance of the global factor for a given country—the larger the forecast error variance attributable to the global factor. Similar interpretation can be given with respect to the other parameters in the expressions.

[1] This is the framework used in Matheson (2013).

Matheson (2013) estimates the dynamic factor model represented by (1) and (2) and provides estimates of both *FEV(G)* and *FEV(R)* for 185 countries from 1990–2011. We will take these values as benchmarks for interdependence among these countries and ask whether *WEO* forecasts incorporate this interdependence.

A natural way to do so would be to estimate a model of the same form as that used by Matheson but to use *WEO* forecasts, rather than actual values, of growth rates as the dependent variable. Unfortunately, doing so is made difficult by the fact that *WEO* forecasts are only available semi-annually for the majority of IMF members, and estimation of the dynamic factor model requires more frequent observations.

For this reason we convert the dynamic factor model of Matheson into a static model by working with forecast revisions rather than the forecasts themselves.

Denote by $FR_t(X_{ij,t+2})$ the *revision* between period *t-1* and period *t* in the forecast for $X_{ij,t+2}$. Then

$$FR_t\left(X_{ij,t+2}\right) = (A_{ij,0}C^2 + A_{ij,1}C)\mu_t^G + (B_{ij,0}D_i^2 + B_{ij,1}D_i)\mu_{it}^R + E_{ij}^2\mu_{ij,t}^I \quad (5)$$

The revision in the forecast will only depend on the current (period *t*) global, regional, and idiosyncratic shocks. The fraction of the variance of forecast revisions accounted for by the global factor is

$$FRV(G) = \frac{1}{1+\frac{D^2(B_0D+B_1)^2\sigma_R^2}{C^2(A_0C+A_1)^2\sigma_G^2}+\frac{E^2\sigma_I^2}{C^2(A_0C+A_1)^2\sigma_G^2}} \quad (6)$$

which, as expected, will depend on the same parameters as the *FEV(G)* in equation (3).

To simplify the relationships between the proportion of forecast *error* variances and the forecast *revision* variance, consider the special case where $A_1=kA_0$, $B_1=kB_0$, and $C=D=E$.

Then we get

$$FEV(G) = \frac{1}{1+\frac{B_0^2\sigma_R^2}{A_0^2\sigma_G^2}+\frac{(C^2+1)\sigma_I^2}{A_0^2(C^2+1+k^2)\sigma_G^2}} \quad (7)$$

$$FEV(R) = \frac{1}{\frac{A_0^2\sigma_G^2}{B_0^2\sigma_R^2}+1+\frac{(C^2+1)\sigma_I^2}{B_0^2(C^2+1+k^2)\sigma_R^2}} \quad (8)$$

and

$$FRV(G) = \frac{1}{1+\frac{B_0^2\sigma_R^2}{A_0^2\sigma_G^2}+\frac{\sigma_I^2}{A_0^2(C+k)^2\sigma_G^2}} \quad (9)$$

Consider now three countries, X_G, X_R, and X_I, which as the subscripts indicate differ by the importance of the global, regional, and idiosyncratic factors. To be concrete, suppose the loading of country *G* on the global factor A^G is large compared to the corresponding loading for countries *R* and *I*. Suppose also that the variance of the regional factor for country *R* is larger than that of countries *G* and *I*, and that finally the variance of the idiosyncratic factor is particularly large for country *I*.

Inspection of the expressions for the forecast error variance decompositions and the forecast revision decomposition would yield the following predictions:

- A country for which *FEV(G)* is high (low) should also have a high (low) *FRV(G)*.
- A country for which *FEV(R)* is high may have a high or low *FRV(G)* depending on the source of the high *FER(R)*, i.e., whether it is due to a high value of σ_R relative to σ_I or a high value of *B*.
- If the value of *FEV* increases over time either because *A* increases relative to *B* or because the variance of the global shocks increases relative to the regional or idiosyncratic shocks then *FRV* should increase over time as well.

We estimate *FRV(G)* and investigate whether the relationships between this measure and the forecast error variances estimated by Matheson conform to these theoretical predictions.

ANNEX 9. SAMPLE USED IN INTERNATIONAL LINKAGES ANALYSIS[1]

Algeria§	Congo§	Hungary§†	Myanmar§	Spain§†‡
Antigua & Barbuda§	Costa Rica§†	India§†	Nepal§	Sri Lanka§†
Argentina§†	Cote D'Ivoire§	Indonesia§†	Netherlands§†‡	St. Kitts and Nevis§
Australia†‡	Croatia†	ran§	New Zealand§†	St. Lucia§
Austria†‡	Cyprus§	Iraq§	Nicaragua§	St. Vincent & the Grenadines§
Azerbaijan†	Czech Republic†	Ireland§†‡	Niger§	Sudan§
Bahamas§	Denmark†‡	Israel§†	Nigeria§†	Suriname§
Bahrain§	Djibouti§	Italy§†‡	Norway§†‡	Swaziland§
Bangladesh§†	Dominica§	Jamaica§	Oman§	Sweden§†‡
Barbados§	Dominican Republic§†	Japan§†‡	Pakistan§†	Switzerland§†‡
Belarus†	Ecuador§†	Jordan§	Panama§†	Taiwan Province of China§†‡
Belgium§†‡	Egypt§†	Kazakhstan†	Papua New Guinea§	Tanzania§
Belize§	El Salvador§	Kenya§	Paraguay§†	Thailand§†
Benin§	Equatorial Guinea§	Korea§†‡	Peru§†	Togo§
Bhutan§	Estonia†	Kuwait§	Philippines§†	Trinidad and Tobago§
Bolivia§†	Ethiopia§	Lao§	Poland§†	Tunisia§
Botswana§	Fiji§	Latvia†	Portugal§†‡	Turkey§†
Brazil§†	Finland§†‡	Lebanon§	Qatar§	Uganda§
Bulgaria§†	France§†‡	Lesotho§	Romania§†	Ukraine†
Burkina Faso§	Gabon§	Liberia§	Russia†	United Arab Emirates§
Burundi§	Gambia§	Lithuania†	Rwanda§	United Kingdom§†‡
Cambodia§	Germany§†‡	Madagascar§	Samoa§	United States§†‡
Cameroon§	Ghana§	Malawi§	Sao Tome & Principe§	Uruguay§†
Canada§†‡	Greece§†	Malaysia§†	Saudi Arabia§†	Uzbekistan†
Cape Verde§	Grenada§	Maldives§	Senegal§	Vanuatu§
Central African Republic§	Guatemala§	Mali§	Seychelles§	Venezuela§†
Chad§	Guinea§	Malta§	Sierra Leone§	Vietnam§†
Chile§†	Guinea-Bissau§	Mauritania§	Singapore§†	Zambia§
China§†	Guyana§	Mauritius§	Slovak Republic†	
Colombia§†	Haiti§	Mexico§†	Slovenia†	
Comoros§	Honduras§	Morocco§	Solomon Islands§	
Congo, DR§	Hong Kong§†	Mozambique§	South Africa§†	

[1] Countries with § are those used in the comparison with Matheson (2013). Countries with † are those used in the Comparisons with Consensus forecasts over time. Countries with ‡ are those used in the comparisons with Consensus forecasts and with Matheson (2013).