IEO
Independent Evaluation Office
of the International Monetary Fund

BACKGROUND PAPER

BP/14/05

IMF Forecasts in the Context of Program Countries

Francesco Luna

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/14/05

IEO Background Paper

Independent Evaluation Office

of the International Monetary Fund

IMF Forecasts in the Context of Program Countries

Prepared by Francesco Luna

February 12, 2014

Abstract

IMF forecasts of GDP growth and inflation for program countries are often perceived as overly optimistic when compared with subsequent out-turns. A review of the relevant literature confirms that this is the conclusion of a number of studies, but reveals some nuances. In particular, econometric analysis suggests that during the evaluation period, 2002–11, the bias was significant only for countries with exceptional access to Fund resources, and that it was generally corrected in the course of the first program review.

Information gathered in interviews and a survey shows that there are several ex ante reasons for any optimistic bias. First, forecasts made in the context of program countries are conditional on the successful implementation of policy measures specified in the program itself. This implies that the forecasts may turn out to be optimistic because the conditions in the program were not all fulfilled. Second, these forecasts are produced in cooperation with the country authorities who, according to some officials interviewed by the evaluation team, typically tend to present a more benign picture to gain popular support for the program. Finally, data quality is often poor in program countries: IMF forecasts made during a crisis period often rely on available data that are eventually revised downward.

Ex Post Assessments of IMF-supported programs are, in general, valuable sources for institutional learning. However, in relation to forecasts they are not well exploited: in practice, their analysis of forecast errors has often been perfunctory.

JEL Classification Numbers: C530

Keywords: IMF-supported programs, forecasting, IMF

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

Contents Page

Abbreviations

BCA	balance of the current account
CPI	consumer price index
DSA	debt sustainability analysis
ECF	Extended Credit Facility
EFF	Extended Fund Facility
EPA	Ex Post Assessment of Members with a Longer-Term Program Engagement
EPE	Ex Post Evaluation of Exceptional Access Arrangements
ESAF	Enhanced Structural Adjustment Facility
ESF	Exogenous Shocks Facility
GAO	Government Accountability Office
GDP	gross domestic product
GGB	general government balance
GRA	General Resources Account
IEO	Independent Evaluation Office
IMF	International Monetary Fund
MONA	Monitoring of Arrangements
PCL	Precautionary Credit Line
PDR	Policy Development and Review Department
PLL	Precautionary Liquidity Line
PRGF	Poverty Reduction and Growth Facility
PRGT	Poverty Reduction and Growth Trust
PSI	policy support instrument
RES	Research Department
RGDP	real gross domestic product
SAF	Structural Adjustment Facility
SBA	Standby Arrangement
SCF	Stand-by Credit Facility
SPR	Strategy, Policy, and Review Department
UN	United Nations
WEO	*World Economic Outlook*

I. INTRODUCTION[1]

1. This paper focuses on IMF short-term forecasts in the context of IMF-supported programs.[2,3] Several considerations motivate this focus. First, more than in other cases, program forecasts have very direct implications for policy decisions. Second, since the forecast incorporated in a program is the result of a negotiation[4] between staff and country authorities, it does not necessarily reflect a purely detached view about the prospects for the economy. Third, these forecasts differ from forecasts associated with regular IMF surveillance since they are conditional on the successful implementation of the policy measures specified in the program.[5] Finally, there is considerable controversy related to the accuracy of such forecasts.

2. The paper addresses three principal questions, all related to program cases:

(i) How do country officials and IMF staff members perceive the quality of IMF forecasts?

(ii) What is the evidence regarding the accuracy of IMF forecasts?

(iii) What learning instruments has the IMF put in place for self assessment of forecast quality? Are they effective?

3. The literature on forecasts in program contexts has dealt mainly with GDP growth and inflation.[6] Although the main emphasis of this paper is on these variables, we devote attention to others as well. Econometric analysis in Section IV covers forecasts of GDP growth, CPI inflation, general government fiscal balance, and the country's external current account. In Section V, which reviews the Fund's self-appraisals of forecasting in program cases, we look at six variables: forecasts of GDP growth, inflation, government balance, external current account, public debt, and external debt.

---

[1] The author is grateful to Hans Genberg, Carlos de Resende, Andrew Martinez, and Franz Loyola for their valuable comments and suggestions. Franz Loyola, furthermore, provided excellent research assistance.

[2] We consider as short-term forecasts those for the current year and for one year ahead. De Resende (2014) analyzes the Fund's medium-term *WEO* forecasts including those for program countries.

[3] For a brief description of an IMF-supported program and the Fund's program-review practices, see Annex 1.

[4] It should be stressed that the word "negotiation" is standard IMF language and summarizes the process of discussion and subsequent review leading to the formalization of the country authorities' adjustment program supported by IMF financing. There is no connotation of quid pro quo in the term employed in this context.

[5] For non-program countries it is typically assumed that established policies will be maintained during the forecast period and that only legislated policy changes will be taken into account in the forecast. For program countries, especially in the case of quantitative targets, the country authorities have a strong vested interest in making those forecasts "come true," and they are in a position and have the means to influence the out-turn.

[6] The accuracy of forecasts of fiscal variables has also been studied extensively, especially by IMF economists.

4. The econometric analysis in this paper employs data on program inceptions and first reviews from the IMF database on Monitoring of Fund Arrangements (MONA). The *World Economic Outlook (WEO)* has been the most common source of data in studies of the accuracy of IMF forecasts, but for the purpose of analyzing program cases MONA contains more detailed information. It collects all projections produced in the context of programs and each program's forecasts are updated at each program review.[7]

5. The paper is organized as follows. Section II employs evaluation survey and interview data to assess the perception of the quality and uses of IMF forecasts in the context of program countries. Section III reviews some of the econometric results found in the literature on forecasts in the context of program countries. This literature is relatively limited, and in large part has been produced inside the IMF or by authors who have worked at the IMF or collaborated with IMF staff. Although the results presented in this literature are highly sample-sensitive and not always consistent, they generally suggest that IMF program forecasts for real GDP growth have an optimistic bias.

6. To assess forecast quality, Section IV uses econometric tests to investigate whether the optimistic bias may be specific to "big" or "high-profile" programs (a result featured in parts of the existing literature), and to assess whether the bias is corrected at the time of a program's first review (as suggested in some interviews). We focus on countries with IMF programs approved between 2002 and 2011.[8] We find that for the full sample of program cases only a weak optimistic bias can be seen in forecasts of real GDP growth, but that for programs providing exceptional access to Fund resources[9] the growth forecasts showed a significant optimistic bias. First program reviews usually led to a correction of this bias. An optimistic bias was typically also present in CPI inflation forecasts, but forecasts of the general government balance and external current account balance consistently erred on the pessimistic side.

---

[7] A detailed description of MONA is in Annex 2. Although the MONA data base is public, the various vintages of macroeconomic forecasts produced for more recent programs (since 2002) are not easily accessible. More precisely, the full database is divided into two periods: 1993–2003 and 2002 to present. The reason behind this distinction is the reclassification and restructuring of several economic variables that occurred in the early 2000s. The 1993–2003 database (available at www.imf.org/external/np/pdr/mona/HistoricalData.aspx) contains data related to each program and its subsequent reviews until the program expired. For macroeconomic data, the 2002–to–present database (available at www.imf.org/external/np/pdr/mona/index.aspx) contains, for each program, only the data related to the most recent review.

[8] Because of the need to check forecasts against out-turns, the most recent programs we consider are those for which one-year-ahead forecasts do not go beyond 2011. We exclude programs that are currently ongoing. Annex 3 lists the 103 programs in our sample.

[9] The IMF can lend amounts above normal limits on a case-by-case basis under its Exceptional Access policy, which entails enhanced scrutiny by the Fund's Executive Board. Exceptional access arrangements comprise access to Fund resources beyond (i) an annual limit of 200 percent of the country's quota; and (ii) a cumulative limit of 600 percent of quota, net of scheduled repurchases. For more details, see IMF Decision No. 14064-(08/18), available at www.imf.org/external/pubs/ft/sd/index.asp?decision=14064-(08/18).

7. Section V looks at the Fund's main instrument of self assessment and learning regarding forecasts in the context of program countries. The guidelines for both these types of documents—Ex Post Assessments of Members with a Longer-Term Program Engagement, and Ex Post Evaluations of Exceptional Access Arrangements—recommend an explicit analysis of forecast accuracy. Of the 42 such documents completed between 2006 and 2013, most carried out their analysis in a manner that produced few suggestions for learning from experience. Section VI concludes.

II. Views from Staff and Country Officials: A Survey and Follow-Up Interviews

8. In this section we report findings from a survey conducted by the IEO to elicit views from country authorities and IMF desk economists about IMF forecasts.[10] We complement these findings with information obtained in follow-up interviews with IMF staff and staff from Executive Directors' offices.

9. Country authorities are generally satisfied with the usefulness and quality of IMF forecasts as well as with the interaction with IMF staff during the forecast process. This is the case for member countries in general but also for countries with IMF programs. Across the membership as a whole, forecasts for advanced countries and for the world are viewed as more useful than forecasts for the official's own country. Among officials in program countries, however, the reverse tends to be the case, albeit with a small margin.

10. With respect to the relative value of short- and medium-term forecasts for policy discussions, officials from program countries tend to place a higher value on the latter than do officials from other countries.

11. Regarding the methods desk economists use to produce forecasts, as well as the reasons for their choice of methods, the survey results are very similar for program and non-program cases: simple spreadsheet analysis based on the IMF's macro framework is by far the most used method, and data availability the most common reason for the choice. Judgment is likewise a very important ingredient in the production of forecasts, somewhat more so in program contexts than elsewhere. Interestingly for the choice of forecast method, departmental guidance tends to be more important in program than in non-program countries.

12. Post-survey follow-up interviews with IMF staff revealed substantial frustration with data availability and quality for low-income countries, especially program countries. Staff frequently cited data shortcomings as the main reason for poor forecast accuracy and for their inability to use more sophisticated forecasting methods.

[10] The survey results are exhaustively described by Genberg and Martinez (2014).

13. Independently of the level of development of the country they work with, numerous staff members indicated that they would like to have more guidance regarding the appropriate forecast method to use.

14. The main findings from the interviews with staff from the offices of Executive Directors representing program countries complement and provide additional nuances to the survey findings.

- In almost all cases, program forecasts were described as the result of a "cooperative process" or of "explicit negotiations" with the IMF team.[11] Typically, such discussions focus on the forecast numbers, but in cases where the capacity and resources of the authorities are sufficiently developed, they also touch on the models used by the IMF team to produce forecasts.

- The importance of the mission chief and the turnover of the IMF team were mentioned several times. In general, the degree of cooperation in the programming process and in reaching agreement on forecasts was seen to depend on the "chemistry" between the IMF team and the authorities and in particular on how friendly and available the mission chief is.

- Interviewees said that typically, the IMF will arrive in the country with initial projections that are more conservative (or less optimistic) than the authorities'. During the mission, the IMF team position loosens and a compromise is found on "more optimistic" ground.

- The value attached to medium-term projections varies widely among respondents; important factors are the level of development of the economy, its institutional strength, and its dependence on natural resources. One extreme of the spectrum of responses is that "medium-term projections should not be available" because the built-in assumption of a rapidly closing output gap offers the authorities a false sense of security or the incentive to overspend. At the other extreme, some resource-rich countries count on such forecasts to discipline government spending.

- A few respondents insisted that the IMF should be more open in admitting its errors in forecasting, especially in the case of program countries.[12]

[11] This is not a new finding. See for example IMF (2011).

[12] On this point it should be noted that many IMF staff interviewed were of the opposite view: (i) they believe that past errors are taken into consideration in new forecasts (even though there is no formal process for determining the accuracy of forecasts), (ii) a formal exercise would be a waste of time since the forecast process is for many low-income countries largely based on judgment and there is no model to improve upon, and (iii) it would be counterproductive unless the analysis of forecast errors were to be made in comparison with other forecasters. The "quality" of the forecasts, they argue, should be ascertained against a benchmark and not in absolute terms.

III. Literature Survey on the Accuracy of Forecasts

15. The econometric literature has investigated the quality of IMF forecasts along three dimensions: bias (do positive and negative errors tend to cancel each other out on average?), efficiency (are forecast errors uncorrelated with variables known at the time of the forecast?), and accuracy (are the forecast errors of the IMF smaller than those of other forecasters or than those in forecasts obtained with naïve methods? Do they correctly predict the direction of change)?

16. Rather few econometric studies have addressed the accuracy of forecasts in the specific context of IMF programs.[13] Without claiming to be exhaustive, this section reports on the more significant findings.

17. In an early study, Goldstein (1986) discusses the global impact of IMF programs and proposes a list of five "measuring rods," one of which is a "normative measure: the difference between performance under the program and the performance specified in its [the program's] targets."[14] From his chosen sample (the set of programs in 1981) Goldstein concludes, first, that programs based on demand-driven stabilization policies seem to have more optimistic forecasts than supply-oriented and mixed-strategy programs, and, second, that considering that the general expectation for a prompt recovery in the world economic activity in 1981 proved overoptimistic, the program forecast errors are quite similar to those recorded in non-program countries.[15]

18. Since the late 1980s the IMF Research Department has commissioned four external evaluations of *WEO* forecasts.[16] The first of these, Artis (1988), makes no explicit mention of projections in the context of program countries. In an update of the first study, Artis (1996) briefly mentions forecasts in the context of IMF programs. He points out that these forecasts are included in the *WEO* and that this fact could explain why *WEO* forecasts for developing countries (many of which have an IMF program) appear to be less accurate than others: "some of the forecasts incorporate data for countries under IMF stabilization programmes, where the programme targets are taken as the forecast."[17]

19. Timmerman (2006), in another commissioned study of *WEO* forecasts, touches marginally on forecasts in the context of program countries: he finds a particularly large bias

[13] On the other hand, a plethora of papers scrutinize some "high-profile" programs arguing that some egregious mistakes were made. See for example Rosnick and Weisbrot (2007) on Argentina and, more recently Aslund (2013) and Sterne (2013a, b) on Greece.

[14] Goldstein (1986) p. 3.

[15] Goldstein concludes that "any underachievement of, say growth targets, need not imply an ineffective program." Goldstein (1986) p. 5.

[16] Freedman (2014) reviews this series of papers.

[17] Artis (1996), p. 34.

in GDP growth projections in program countries and suggests that the bias could be at least reduced by giving more appropriate weight to international linkages. He notes that U.S. and German GDP growth forecasts are correlated with forecast errors for many program countries.

20. The most recent of the commissioned studies (Faust, 2013) confirms the optimistic bias for program countries reported in previous evaluations. Faust suggests that the *WEO* should explicitly treat such forecasts as conditional on successful implementation of specific policies detailed in the program. He goes further, to propose that the preparation of unconditional forecasts be contracted to some external forecaster.

21. Also interesting from our perspective is Faust's emphasis on the ever-changing nature of the data-generating process the economic systems that the forecaster is trying to depict—which requires the forecaster continuously to update his "model" of the economy. Faust points out that traditional methods of evaluating forecasts may not be adequate in such a circumstance. These considerations are particularly relevant for countries that are very likely undergoing structural change triggered or magnified by a crisis.

22. Musso and Phillips (2002) analyze IMF programs negotiated between 1993 and 1997. They look at bias, efficiency, and accuracy in forecasts of five variables: GDP growth, inflation, and three balance of payments measures (current account balance, capital account balance, and changes in official reserves). They find that forecasts of inflation and foreign reserves are systematically below out-turns but that there is no statistical bias in forecasts of GDP growth or of current and capital account balances. They note, though, that the results may differ for "big" programs; specifically, GDP growth forecasts in these cases show an optimistic bias. Examining efficiency, they find that GDP growth forecasts pass several tests, but that for the other variables the forecasts do not appear to encompass all available information. In terms of accuracy, they find that only the program forecasts for GDP growth, inflation rates, and changes in official reserves perform well above the naïve benchmark (a random walk).

23. Golosov and King (2002) concentrate on tax-revenue forecasts for a sample of low-income program countries between 1993 and 1999. They find that forecast accuracy is poor and that there is a positive bias in forecasts of the ratio of tax revenue to GDP, though not in forecasts of the percentage change in nominal tax revenue.[18] The authors ask whether the factor that is typically held responsible for ex post bias in forecasts—poor implementation of program measures—can explain the biases in their sample. Interestingly, although their results confirm the conditional nature of program forecasts, they find "quite weak" statistical evidence that compliance with program conditions translates into smaller bias; in fact, this bias is present even for programs that achieved normal completion. The authors conclude that additional reasons must be behind the bias. They show that when tax revenues are an explicit target, the bias is smaller but still present; that there is no significant geographical factor

[18] There is a positive correlation between the forecast errors in tax revenues as a percentage of GDP and nominal GDP growth forecast errors for program countries. Errors tend to offset each other.

(justified in terms of possibly different economic structures); and that the bias is larger for programs that are interrupted and "for programs with a longer forecast horizon."[19]

24. In 2003, the U.S. Government Accountability Office (GAO), prompted by U.S. Congressional concerns regarding the accuracy of IMF forecasts, completed an analysis of the *WEO* forecasts. The GAO results show that *WEO* forecasts of GDP growth and inflation in general show an optimistic bias. Forecasts for developing countries that were or had been in a program with IMF performed better than forecasts for countries that had never had an IMF program. In the report's words, this may have been because the forecasts for program countries "are produced under conditions of greater staff scrutiny."[20]

25. Pursuing the multiple factors behind program forecast bias, Atoyan and others (2004) and Atoyan and Conway (2011) assess the impact on the bias caused by (i) poor data, (ii) model specification, (iii) poor policy implementation, and (iv) random errors. They focus on forecasts of fiscal balance and external current account balance and employ data from the MONA database for 145 programs between 1993 and 2001. In addition, they study the effect of programs' first-review updates on the bias, efficiency, and accuracy of forecasts. Their results suggest that a large part of the bias in the forecasts for their chosen macro variables is due to the model[21] employed by the IMF. The alternative used by the authors is a vector auto regression (VAR). However, they find that the poor measurement of initial conditions is quite important, and actually the main cause of error in current account forecasts. Exogenous errors account for between 30 percent (for fiscal balance) and 40 percent (for current account) of the total projection error. The policy implementation performance of programs does not appear to play a fundamental role. Forecasts made at the time of a program's first review are found to be superior, mainly in terms of bias and accuracy, to forecasts made at program inception; the authors conclude that this improvement is to be ascribed to better data and to learning-by-doing at the modeling stage.

26. Baqir, Ramcharan, and Sahay (2005), based on a sample of 94 programs between 1989 and 2002, find an optimistic bias in forecasts of GDP growth and inflation, but no bias in forecasts of the current account. The bias in forecasts of GDP and inflation is found to persist after controlling for shocks and policy implementation (although these authors do not consider data quality, as suggested by Atoyan and others, 2004). Baqir and others point to the nature of the arrangement as an element correlated with the magnitude of the bias and—like Musso and Phillips (2002)—they note that program size may also play a role: "growth projections are more optimistic in stand-by arrangements (SBAs) than in Poverty Reduction and Growth Facility programs (PRGFs), with one caveat: the projections in the high-profile SBAs were more realistic than in other SBAs and PRGFs, although the direction of the bias

[19] Golosov and King (2002), p. 19.

[20] U.S. GAO (2003), p. 39.

[21] "Model" has to be interpreted in a broad sense to encompass the set of models employed as well as the role played by judgment.

was the same in all types of program."[22] Similarly, inflation forecasts appear optimistic across the sample, but those in "high-profile" stand-by arrangements are less biased than average. This result appears at odds with the common perception that forecasts for countries with big programs are on average less accurate. The specific "high-profile" programs considered in their study are: Argentina, Brazil, Indonesia, the Republic of Korea, Mexico, the Russian Federation, Thailand, Turkey, and Uruguay.

27. At the end of 2005 the IMF Policy Development and Review Department (PDR) prepared an evaluation of debt projections in the context of a general debt sustainability analysis (DSA) template review. The PDR study confirmed a significant small optimistic bias—of 1 percentage point to 2 percentage points of GDP—for the overall sample of 136 countries. However, it found no systematic evidence of a larger forecast bias for program countries than for surveillance countries.

28. As this brief review suggests, the evidence on the general quality of forecasts in the context of IMF-supported programs is mixed but there is a tendency to find an optimistic bias in GDP growth forecasts. The following arguments summarize the most common explanations proposed in the literature for this optimism. These arguments also found support in the evaluation team's interviews with IMF staff, representatives from Executive Directors' offices, and country officials:

- Projections may be aimed at influencing program outcomes. The desire to trigger specific behavior in economic agents or, at a minimum, to generate a "morale effect" would create an incentive to err on the optimistic side. Taking the argument to the extreme, interviewees—both staff and country officials—recognized that the IMF cannot risk being considered the "crisis catalyst" by presenting too bleak a picture of an already weak economy.[23]

- Projections are the results of a cooperative process that will guarantee the country authorities' ownership. Interviews with IMF staff and country officials suggested that there is a "public relations/marketing" bias: a program's projections should be optimistic to be accepted more easily both by the IMF Executive Board and the parliament/public opinion in the country, to counterbalance the stigma that sometimes is attached to having a Fund program. In particular, if the authorities are leaning towards very optimistic projections, the final result of the process will be biased in that direction.[24]

---

[22] Baqir and others (2005), p. 270. Their paper uses "high profile" to identify "large access" programs with lending exceeding 2 billion SDRs.

[23] This point was raised recently in Sterne (2013a).

[24] In some cases even a pessimistic bias serves some political agenda. The Ex Post Assessment for Argentina (IMF, 2006a) suggests that a pessimistic bias on GDP growth played in favor of Argentina's pleas for a favorable debt restructuring.

- As noted in the literature, and confirmed in the interviews, poor data may greatly hinder the quality of forecasts in general and program projections in particular.[25]

IV. Bias in Program Country Forecasts: "Big" Programs and the First Review's Impact

29. In this section we address two series of hypotheses that are implicit in the literature and have emerged from the interviews held for this evaluation. First, we study differences in the forecast bias between a program's inception and its first review. Looking at GDP growth and inflation forecasts, our hypothesis is that the optimistic bias—recorded in several empirical studies for program countries—should be largely corrected at the time of the first review, occurring three to six months after the beginning of the program.

30. Following up on the reasons adduced in the literature and in the interviews,[26] we test whether the optimistic bias disappears once the program is in place—that is, once the program has been approved by the country's authorities, accepted by public opinion, and signed off by the IMF Executive Board. Pursuing the argument that emerged in the interviews with staff and country officials one step further, the forecasts of those variables more strictly linked to the program's quantitative targets should show a persistent or even reinforced bias so as to make the forecast targets easier to meet.

31. Second, we address the role of "big" or "high-profile" programs and the nature of the arrangements with the Fund, distinguishing between concessional or Poverty Reduction and Growth Trust programs (Extended Credit Facility, Extended Structural Adjustment Fund, Exogenous Shocks Facility, Poverty Reduction and Growth Facility, Policy Support Instrument, and Stand-by Credit Facility) and non-concessional or General Resources Account programs (Stand-by arrangements, Extended Fund Facility, Precautionary Credit Line, and Precautionary Liquidity Line).

32. As noted in Section I, we extract our information from the MONA database for the period 2002–11.[27] For the103 programs that were approved in this period (listed in Annex 3), we investigate the following hypotheses: (i) whether an optimistic bias is mainly characteristic of "big" programs, and (ii) whether the bias will largely be corrected or reinforced at the time of the first review. We try different ways to capture the idea of "big"

[25]At the beginning of an economic downturn, poor collection and aggregation of partial data tend to translate in estimated data above their actual level, but this is reflected only in subsequent revisions. An overstated starting point distorts upward especially GDP growth forecasts.

[26] These reasons are reported in Paragraph 28.

[27] Because of the need to check forecasts against out-turns, the most recent programs we consider are those for which one-year-ahead forecasts do not go beyond 2011. We exclude programs that are currently ongoing. Annex 3 contains a list of the programs in our sample.

programs. First, we refer to exceptional access programs[28] as our "big" programs. Then, following the idea of "high-profile" programs introduced by Baqir and others (2005), we look at arrangements disbursing more than two billion SDRs. Finally, we also try to ascertain whether the nature of the arrangement—concessional or market-rate based—has an impact on the bias.

33. We study the forecast error properties for four variables: GDP growth, CPI inflation rate, general government balance,[29] and current account balance. We look at two horizons: current year and one year ahead. Note that, for each arrangement, we employ two observations per horizon period: the one at inception and the one at first review. In all cases, we focus on the forecast error. For each variable and forecasting horizon, we present the results of five regressions:

$$\varepsilon_t = \beta_0 + \beta_{rev}\, rev + \eta_t \quad (1)$$

$$\varepsilon_t = \beta_0 + \beta_{rev}\, rev + \beta_{same-yr}\, S\_year + \beta_{rev-sY} rev * S\_year + \eta_t \quad (2)$$

$$\varepsilon_t = \beta_0 + \beta_{rev}\, rev + \beta_{till-hor}\, till\_horizon + \beta_{EA} EA + \beta_{rev-EA} rev * EA + \eta_t \quad (3)$$

$$\varepsilon_t = \beta_0 + \beta_{rev}\, rev + \beta_{till-hor}\, till\_horizon + \beta_{Big} BIG + \beta_{rev-bg} rev * BIG + \eta_t \quad (4)$$

$$\varepsilon_t = \beta_0 + \beta_{rev}\, rev + \beta_{till-hor}\, till\_horizon + \beta_{GRA} GRA + \beta_{rev-GRA} rev * GRA + \eta_t \quad (5)$$

where ε_t, the forecast error, is defined as actual value[30] minus forecast value; β_0 is the coefficient for the constant, *rev* is a dummy variable that is equal to one when the forecast is expressed at first review, S_year is a dummy that is equal to one if the first review occurs in the same calendar year as program inception; $till\_horizon$ is the number of weeks between the forecast formulation and the forecast horizon (either end of current year or one year ahead); *EA* is a dummy that takes the value one if the program considered is "exceptional access." In addition, rev **EA* is the interaction dummy (which will be one when both rev and *EA* are true), *BIG* is a dummy that identifies the arrangements with disbursement above two billion SDRs, and *GRA* is a dummy that takes the value one where the arrangement is non-concessional (EFF, PCL, PLL, or SBA), and zero where the arrangement is concessional (ECF, ESAF, ESF, PRGF, PSI, or SCF).

[28] Exceptional Access arrangements comprise access beyond (i) an annual limit of 200 percent of quota; and (ii) a cumulative limit of 600 percent of quota, net of scheduled repurchases. For more details, refer to IMF Decision No. 14064-(08/18) available at www.imf.org/external/pubs/ft/sd/index.asp?decision=14064-(08/18). We will use this type of arrangements to identify "objectively" big programs.

[29] According to the evaluation survey, these first three are the forecast variables that country authorities consider the most useful (Genberg and Martinez, 2014).

[30] "Actual" is defined as the out-turn of each variable as recorded two years after the time the forecast refers to. Because of the interest in the more recent IMF programs, we have used the latest *WEO* publication (April 2013) for the 2011 actual out-turns, even though less than two years has passed since end–2011.

34. Table 1 helps interpret the results of the panel regressions we present subsequently:

Table 1. Interpreting the signs of the coefficients (errors = actual - forecast)

Variable	Coefficient sign	Interpretation
RGDP	$\beta_0 < 0$	Optimistic
	$\beta_0 > 0$	Pessimistic
	$\beta_{rev} < 0$ & $\beta_0 < 0$	Worsening of Bias
	$\beta_{rev} < 0$ & $\beta_0 > 0$	Correction of Bias
	$\beta_{rev} > 0$ & $\beta_0 < 0$	Correction of Bias
	$\beta_{rev} > 0$ & $\beta_0 > 0$	Worsening of Bias
Inflation	$\beta_0 < 0$	Pessimistic
	$\beta_0 > 0$	Optimistic
	$\beta_{rev} < 0$ & $\beta_0 < 0$	Worsening of Bias
	$\beta_{rev} < 0$ & $\beta_0 > 0$	Correction of Bias
	$\beta_{rev} > 0$ & $\beta_0 < 0$	Correction of Bias
	$\beta_{rev} > 0$ & $\beta_0 > 0$	Worsening of Bias
General Gov Balance[1]	$\beta_0 < 0$	Optimistic
	$\beta_0 > 0$	Pessimistic
	$\beta_{rev} < 0$ & $\beta_0 < 0$	Worsening of Bias
	$\beta_{rev} < 0$ & $\beta_0 > 0$	Correction of Bias
	$\beta_{rev} > 0$ & $\beta_0 < 0$	Correction of Bias
	$\beta_{rev} > 0$ & $\beta_0 > 0$	Worsening of Bias
Current Account Balance	$\beta_0 < 0$	Optimistic
	$\beta_0 > 0$	Pessimistic
	$\beta_{rev} < 0$ & $\beta_0 < 0$	Worsening of Bias
	$\beta_{rev} < 0$ & $\beta_0 > 0$	Correction of Bias
	$\beta_{rev} > 0$ & $\beta_0 < 0$	Correction of Bias
	$\beta_{rev} > 0$ & $\beta_0 > 0$	Worsening of Bias

[1] For general government balance and current account balance we adopt the same convention used for GDP growth. Hence, without implying any normative connotation, a forecast budget deficit smaller (or a surplus larger) than the out-turn is considered "optimistic."

35. Table 2 records the regression results corresponding to equation (1) for all variables. Real GDP growth and CPI inflation coefficients are expressed in percentage points, while government and current account balances are expressed in percentage points of GDP. For example, the coefficient attached to the constant (the bias in the real GDP regression) is equal to -0.242, which means that on average the real GDP growth rate turns out to be 0.242 percentage points less than projected. By contrast, the current account balance on average turns out to be 1.52 percent of GDP larger than projected.

- *Real GDP* (*RGDP*). This initial result for both horizons goes somewhat against the general perception and a large part of the literature on the topic. For the overall sample (206 observations from 103 different programs), there appears to be weak evidence of an optimistic bias at program inception (the coefficient is negative,

indicating optimistic forecasts, but it is not statistically significant). However, there is evidence that the first review imposes a correction so that the bias for the full sample at the time of first review is pessimistic.

- *CPI Inflation (PCPI)*. In the case of inflation, there is evidence of an optimistic bias in the current-year forecasts, and the first review imposes a significant correction so that the bias at first review is much reduced (the coefficient is still positive but non-significant). In the one-year-ahead forecasts there is significant evidence of an optimistic bias at first review, and indeed the specific first-review effect, although statistically non-significant, is to reinforce the bias.

- *General Government Balance (GGB).* For the current-year forecast, there is a weak pessimistic bias at program inception and the first review imposes a significant correction, so that the bias switches to optimistic (non-significant). There is quite a significant pessimistic bias for the one-year forecast, and the review imposes a significant correction that renders the bias non-significant.

- *Current Account Balance (BCA).* A similar story holds for the current account balance: at program inception there is some weak pessimistic bias in the forecasts for the current-year and for one-year-ahead, that is corrected with the first review. However, neither bias is statistically significant.

Table 2. Regression results corresponding to equation (1)

	Current year forecast				One year ahead forecast			
Variables	RGDP	PCPI	GGB	BCA	RGDP	PCPI	GGB	BCA
β_{rev}	0.611*	-0.698**	-1.714*	-1.043	0.494	1.112	-0.748*	-0.709
	(0.368)	(0.306)	(0.933)	(0.914)	(0.478)	(1.175)	(0.403)	(1.050)
β_0	-0.242	1.222**	0.873	1.520	-0.336	0.426	3.646**	0.962
	(0.522)	(0.589)	(1.064)	(1.082)	(0.584)	(1.235)	(1.813)	(1.424)
Observations	206	206	158	106	206	206	158	106
Arrangements	103	103	79	53	103	103	79	53
Rho	0.664	0.819	0.693	0.531	0.464	0.193	0.976	0.742
Wald Prob > chi2	0.0972	0.0223	0.0663	0.254	0.301	0.344	0.0634	0.499
B of first review								
(β_0 + β_{rev})	0.369	0.523	-0.840	0.476	0.158	1.539	2.897	0.253
Standard Error	0.363	0.412	1.305	0.781	0.291	0.434	1.864	1.499
p-value	0.310	0.204	0.520	0.542	0.587	0.000391	0.120	0.866

Robust standard errors in parentheses.

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

36. All in all, there is some support for the perception that forecasts are optimistic, especially for GDP growth and inflation, although it is not always very strong in this data sample. On the other hand, the corrective role of first reviews is largely confirmed.

37. The corrective tendency of first reviews might simply be the effect of a growing information set that renders the forecasts (for a now closer horizon) more precise. To try to separate the passage of time from the impact that other considerations may dictate, we introduce in equation (2) the same-year dummy. The idea is that if the first review maintains some corrective power after controlling for time passed, then, indeed, there is more to first reviews than a simple update of the information set. In Table 3, we construct a "heat map" to help interpret the results of the regression. The shades of red identify optimistic biases. Pink stands for a statistical non-significant optimistic bias. We use orange for an optimistic bias with a significance level between 10 percent and 1 percent, and red for very significant optimistic biases. Similarly, the deepening shades of blue indicate pessimistic biases of rising statistical significance.

- *RGDP.* The time factor appears to contribute to the corrective impact of first reviews. Note that for the one-year forecasts, the correction leads into the pessimistic territory (although the result is statistically not significant).

- *PCPI.* The correction introduced by time is only marginal (and statistically insignificant). However, in the case of same-year reviews, for the one-year-ahead forecasts, the review introduces a boost of optimism that makes the bias optimistic and very significant.

- *GGB.* The pessimistic bias is evident for both horizons and is quite strong (and significant) for the current year; the same-year-review dummy reinforces the negative bias in the case of current-year forecasts, but contributes to the correction in the one-year-ahead forecasts along with the first review dummy.

- *BCA.* As for Equation 1, the pessimistic bias recorded at program inception is, for both horizons, corrected in the first review. Since the same-year dummy is unable to explain the full correction, we infer that the first-review dummy captures other considerations affecting the review process.

38. The conclusion that can be drawn from Table 3 is that the sometimes significant improvement in accuracy between inception and first review can be explained only in part by the natural increase in the information set over time, and that the first review itself plays a corrective role in certain cases.

Table 3. Regression results corresponding to equation (2)

	Current year forecast				One year ahead forecast			
Variables	RGDP	PCPI	GGB	BCA	RGDP	PCPI	GGB	BCA
$\beta_{same\text{-}Yr}$	0.124	-1.283	2.386	-0.602	1.181	-2.589	-3.022	-1.760
	(1.132)	(1.279)	(2.510)	(2.237)	(1.301)	(2.285)	(4.206)	(2.817)
β_{rev}	0.791	-0.752	-2.740	-2.129	1.064	-0.381	-0.543	-1.211
	(0.761)	(0.503)	(2.169)	(1.440)	(1.072)	(0.266)	(0.573)	(1.710)
$\beta_{rev\text{-}sY}$	-0.320	0.0965	1.762	2.054	-1.012	2.698	-0.353	0.951
	(0.814)	(0.629)	(2.204)	(1.843)	(1.085)	(2.127)	(0.802)	(2.145)
β_0	-0.312	1.932	-0.516	1.838	-1.001	1.859**	5.405	1.891
	(1.021)	(1.202)	(2.475)	(1.901)	(1.243)	(0.841)	(4.048)	(1.701)
Observations	206	206	158	106	206	206	158	106
No. of Arrangements	103	103	79	53	103	103	79	53
Rho	0.665	0.817	0.688	0.539	0.466	0.196	0.976	0.742
Wald Prob > chi2	0.286	0.132	0.0325	0.384	0.775	0.348	0.187	0.831
Coefficient								
β_0	-0.312	1.932	-0.516	1.838	-1.001	1.859	5.405	1.891
Standard Error	1.021	1.202	2.475	1.901	1.243	0.841	4.048	1.701
p-value	0.760	0.108	0.835	0.334	0.421	0.0271	0.182	0.266
$\beta_0 + \beta_{same\text{-}Yr}$	-0.188	0.649	1.870	1.236	0.180	-0.730	2.383	0.131
Standard Error	0.490	0.437	0.419	1.180	0.383	2.124	1.142	2.246
p-value	0.701	0.137	8.03e-06	0.295	0.639	0.731	0.0370	0.953
$\beta_0 + \beta_{rev}$	0.479	1.180	-3.256	-0.291	0.0633	1.478	4.863	0.680
Standard Error	0.707	0.862	3.039	1.272	0.491	0.862	4.225	2.030
p-value	0.497	0.171	0.284	0.819	0.897	0.0866	0.250	0.738
$\beta_0 + \beta_{rev} + \beta_{same\text{-}Yr} + \beta_{rev\text{-}sY}$	0.283	-0.00656	0.893	1.161	0.232	1.588	1.487	-0.129
Standard Error	0.345	0.252	0.418	0.949	0.354	0.369	1.042	2.215
p-value	0.412	0.979	0.0326	0.221	0.512	1.70e-05	0.153	0.954
First Review Impact								
β_{rev}	0.791	-0.752	-2.740	-2.129	1.064	-0.381	-0.543	-1.211
Standard Error	0.761	0.503	2.169	1.440	1.072	0.266	0.573	1.710
p-value	0.299	0.135	0.207	0.139	0.321	0.153	0.343	0.479
$\beta_{rev} + \beta_{rev\text{-}sY}$	0.471	-0.655	-0.978	-0.0746	0.0519	2.318	-0.896	-0.260
Standard Error	0.287	0.379	0.391	1.150	0.165	2.110	0.562	1.296
p-value	0.101	0.0836	0.0123	0.948	0.753	0.272	0.111	0.841

Pessimistic p<0.01
Pessimistic 0.01<p<0.1
Pessimistic non significant
Optimistic non significant
Optimistic 0.01<p<0.1
Optimistic p<0.01

39. Pursuing this aspect, we check whether the program forecasts are efficient in including all information available to other forecasters, as encompassed in Consensus forecasts.[31] To do this we simply regress the IMF forecast error on the Consensus forecast as:

$$\varepsilon_{IMF:t} = c + \beta F_{con:t}x_t + \eta_t \qquad (4)$$

Where $\varepsilon_{IMF:t}$ is the IMF forecast error, and $F_{con:t}x_t$ is the Consensus forecast of variable x_t published during the same month as (or one month before) the IMF forecast. If the parameter β attached to the Consensus forecast is significant, then it can be concluded that there is some information that was available to the IMF that was not taken into consideration.[32]

40. Table 4 summarizes the results for the available variables: GDP growth and CPI inflation. We again look at two horizons: current year and one year ahead.

Table 4. IMF forecast errors

Consensus Forecasts	RGDPe_0	RGDPe_1	PCPIe_0	PCPIe_1
cons_rgdpf_0	-0.482			
	(0.308)			
cons_rgdpf_1		0.921**		
		(0.404)		
cons_pcpif_0			0.330***	
			(0.118)	
cons_pcpif_1				-0.546
				(1.516)
Constant	1.165*	-2.230*	-2.769***	3.383
	(0.681)	(1.241)	(0.999)	(10.02)
Observations	68	68	68	68
R-squared	0.181	0.460	0.304	0.003
Arrangements	34	34	34	34

Robust standard errors in parentheses.

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

41. The results indicate that there is some information underlying the Consensus forecasts that was not exploited in the IMF forecasts. This may be because the two information sets are not identical. If that is the case, we would expect to find that IMF forecasts could help reduce Consensus forecast errors. We run the reverse regression:

$$\varepsilon_{con:t} = c + \beta F_{IMF:t}x_t + \eta_t$$

[31] Consensus forecasts are published by Consensus Economics, at www.consensuseconomics.com.

[32] For a group of forecasters, another way to address the issue of whether updates take into consideration all information available at the time is to test for "informational rigidities" as suggested by Loungani and others (2011). Unfortunately, the procedure designed by Coibion and Gorodnichenko (2012) is not appropriate in the case of a single forecaster.

42. Indeed, we find that IMF forecasts could have helped reduce Consensus forecast errors (Table 5). Thus, overall the results suggest that the IMF forecasting process adds significant value, but that it may still benefit from a deeper understanding of the factors that underlie Consensus forecasts.

Table 5. Consensus forecast errors

IMF Forecasts	cons_rgdpe_0	cons_rgdpe_1	cons_pcpie_0	cons_pcpie_1
RGDPf_0	-0.223			
	(0.272)			
RGDPf_1		0.570***		
		(0.148)		
PCPIf_0			0.916*	
			(0.464)	
PCPIf_1				-0.00359
				(0.00954)
Constant	0.310	-1.435***	-7.730*	1.090***
	(0.541)	(0.357)	(3.885)	(0.0752)
Observations	68	68	68	68
R-squared	0.060	0.515	0.324	0.001
Arrangements	34	34	34	34

Robust standard errors in parentheses.

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

43. The next issue we investigate is the role of "big" programs, as first suggested by Musso and Phillips (2002) and revisited by Baqir and others (2005). Note that the conclusions drawn by these authors differ: Musso and Phillips conclude that big programs show a larger bias, while Baqir and others conclude that big programs are less biased than the average stand-by agreement.[33] In Table 6 below, regression results based on equation (3) are presented. We equate "big" programs with exceptional access programs and introduce a dummy and the interaction term for first review for exceptional access programs. It is important to stress that in our sample fewer than 15 percent of all IMF programs qualify as "exceptional access" but account for around 85 percent of the total program disbursements. To simplify the presentation of the results, we substitute the "same-year" dummy, with a "time" variable that measures the number of weeks between the time at which the forecast was formulated and its horizon.

44. Looking at the table by column we see, again, no strong evidence of a generalized optimism in IMF program forecasts. However, looking at forecasts of real GDP, the findings suggest a rather strong optimistic bias (for current-year projections) at program inception for exceptional access programs. Such a bias was promptly corrected at first review (especially for exceptional access programs). The one-year-ahead forecasts show a non-significant optimistic bias for programs and, as usual, first reviews introduced some correction.

[33] These two papers use different samples and their set of "high-profile" programs is rather limited (Just 10 instances for Baqir and others). The results in general in the literature appear quite sample-dependent.

Table 6. Regression results corresponding to equation (3)

	Current year forecast				One-year-ahead forecast			
Variables	RGDP	PCPI	GGB	BCA	RGDP	PCPI	GGB	BCA
$\beta_{till\text{-}hor}$	-0.0274*	-0.0128	-0.0388	-0.0204	0.0154	0.0443	-0.0279	-0.00331
	(0.0164)	(0.0156)	(0.0395)	(0.0349)	(0.0225)	(0.0316)	(0.0247)	(0.0362)
β_{rev}	0.245	-0.885***	-1.836	-1.038	0.459	1.075	-0.704*	-0.591
	(0.381)	(0.325)	(1.153)	(1.091)	(0.536)	(1.320)	(0.416)	(1.296)
$\beta_{E.A}$	-3.783***	-2.796**	2.457	1.074	-0.0115	0.638	-4.027*	-1.067
	(1.388)	(1.121)	(1.805)	(2.249)	(1.183)	(1.758)	(2.399)	(2.622)
$\beta_{rev\text{-}E.A}$	2.619***	1.451*	1.422	0.173	0.0816	-0.316	0.155	-0.535
	(0.979)	(0.822)	(1.435)	(1.647)	(0.694)	(1.535)	(0.821)	(1.796)
β_0	0.967	1.919**	1.355*	1.748	-0.684	-0.667	4.874**	1.256
	(0.671)	(0.806)	(0.804)	(1.471)	(1.083)	(2.062)	(2.301)	(1.779)
Observations	206	206	158	106	206	206	158	106
Arrangements	103	103	79	53	103	103	79	53
Rho	0.671	0.820	0.681	0.530	0.467	0.201	0.976	0.740
Wald Prob > chi2	0.00406	0.0322	0.174	0.775	0.577	0.475	0.0831	0.807
Coefficient								
β_0	0.967	1.919	1.355	1.748	-0.684	-0.667	4.874	1.256
Standard Error	0.671	0.806	0.804	1.471	1.083	2.062	2.301	1.779
p-value	0.149	0.0173	0.0919	0.235	0.528	0.746	0.0342	0.480
$\beta_0 + \beta_{E.A}$	-2.816	-0.877	3.812	2.822	-0.696	-0.0295	0.847	0.189
Standard Error	1.332	1.061	1.642	2.126	1.071	1.277	1.353	2.126
p-value	0.0344	0.408	0.0203	0.184	0.516	0.982	0.531	0.929
$\beta_0 + \beta_{rev}$	1.212	1.034	-0.481	0.710	-0.226	0.407	4.170	0.665
Standard Error	0.488	0.639	1.464	1.004	0.610	0.898	2.321	1.855
p-value	0.0130	0.105	0.743	0.480	0.712	0.650	0.0725	0.720
$\beta_0+\beta_{rev}+\beta_{E.A}+\beta_{rev\text{-}E.A}$	0.0477	-0.312	3.399	1.957	-0.155	0.730	0.297	-0.938
Standard Error	0.833	0.529	1.520	1.380	1.123	1.158	0.987	1.797
p-value	0.954	0.555	0.0254	0.156	0.890	0.529	0.763	0.602
First Review Impact								
β_{rev}	0.245	-0.885	-1.836	-1.038	0.459	1.075	-0.704	-0.591
Standard Error	0.381	0.325	1.153	1.091	0.536	1.320	0.416	1.296
p-value	0.520	0.00644	0.111	0.341	0.392	0.415	0.0904	0.648
$\beta_{rev} + \beta_{rev\text{-}E.A}$	2.864	0.565	-0.414	-0.865	0.540	0.759	-0.550	-1.127
Standard Error	0.899	0.744	0.771	1.235	0.401	0.695	0.697	1.236
p-value	0.00144	0.447	0.592	0.484	0.178	0.275	0.431	0.362

Robust standard errors in parentheses.

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

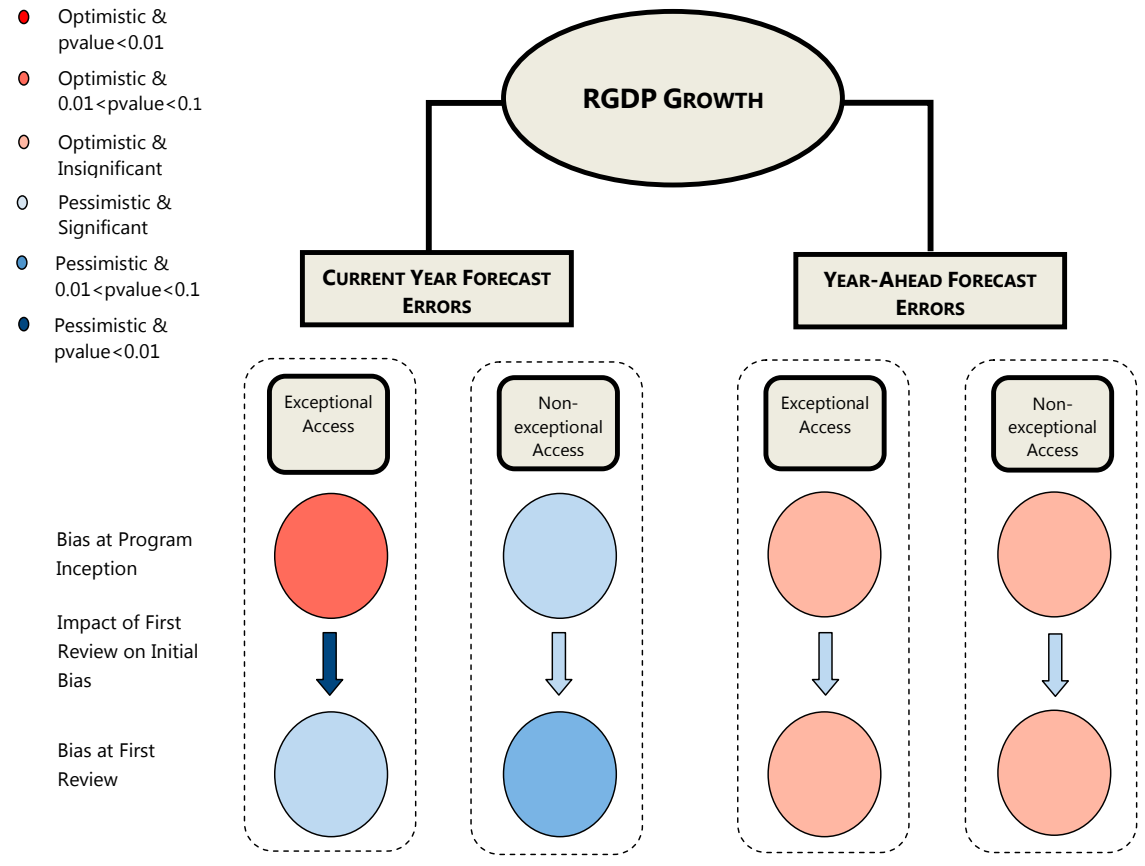
Pessimistic p<0.01
Pessimistic 0.01<p<0.1
Pessimistic non significant
Optimistic non significant
Optimistic 0.01<p<0.1
Optimistic p<0.01

45. There is some trace of an optimistic bias also in the case of CPI inflation projections, especially for the one-year-ahead horizon. This bias was reinforced rather than corrected with the first review for all programs.

46. For the general government balance and the external current account balance, the tendency seems to have been towards a pessimistic bias, especially for government balance at program inception.

47. Figure 1 summarizes the findings for GDP growth. The circles represent the bias identified by the regression at inception and first review, while the arrows indicate the impact of the first review so as to clarify whether the review imposed a correction or, rather, a reinforcement of the initial bias. The color code is the same as described above, with warmer colors indicating increasingly significant levels of optimism. For example, for exceptional access programs, the orange circle on the left represents a statistically significant optimistic bias in GDP growth current-year forecasts. The light-blue circle just below represents the comparable bias at first review. The optimistic bias has disappeared, and the dark blue arrow summarizes the direction and strength of the corrective impact that occurred at the time of the first review.

Figure 1. Forecast errors at program inception and first review



48. These findings are consistent with some of the answers collected in interviews with IMF staff and Executive Directors' offices. In particular, we raised the issue of the implicit contradiction between an optimistic GDP forecast and fiscal targets—designed to be ambitious, but feasible—that should be linked with GDP dynamics: higher GDP growth would suggest harder-to-meet fiscal targets. The answer we received was that GDP growth and fiscal targets are not tightly linked. First, authorities tend to keep some room for maneuver on the revenue and expenditure side so as to meet the balance target if projected revenues fall short or if unexpected expenditures arise. Second, lower than expected GDP growth offers the authorities a very good rationale (outside of the authorities' responsibility) for why fiscal targets could not be met, in case a waiver is needed.

49. Furthermore, IMF staff explained the persistent pessimistic bias in forecasts of government fiscal balance and current account balance as the result of the cooperative process in program design. The pessimistic bias implies that some of the most common IMF program targets—fiscal deficit and foreign reserves—will be easier to meet.

50. The results are suggestive as to why IMF forecasts are often perceived to have an optimistic bias. Indeed, it is mainly high-profile programs that receive particular attention from the media. Hence, we follow a different approach to try to determine whether such a perception is driven by the poor performance of forecasts produced in the context of exceptional access programs. For GDP growth projections, we take as benchmark the Consensus forecasts when available. Figure 9 shows the errors of IMF forecasts for exceptional access program countries and Consensus forecasts. The top two panels refer to the current-year forecast errors at program inception and at first review, respectively.

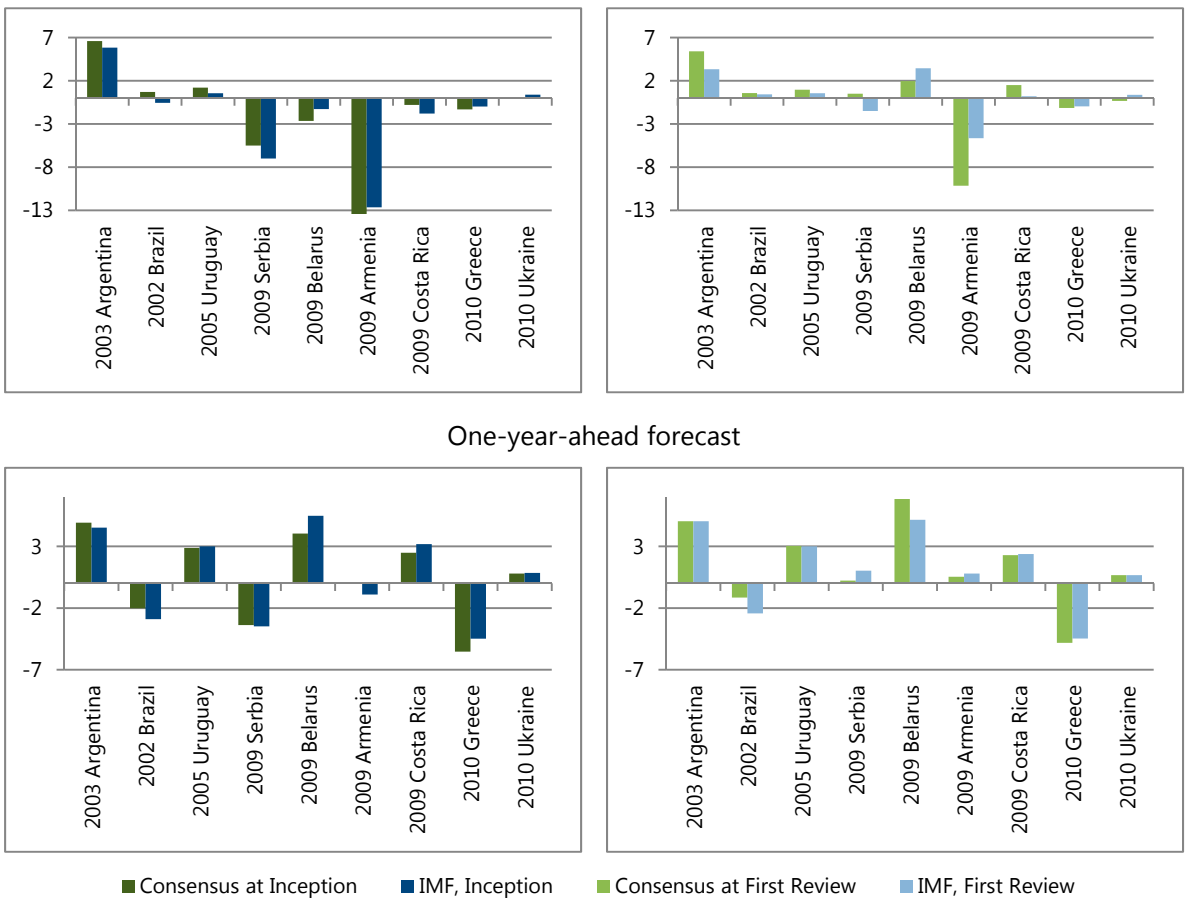
51. At program inception, the IMF GDP forecasts for four countries (out of nine considered) were more optimistic than Consensus forecasts (Figure 2). However, it is important to stress that such errors were almost invariably corrected at the time of first review, whereas the forecast errors for Consensus forecasts were quite "sticky."[34] In making one-year-ahead forecasts, IMF fared significantly better than Consensus even at program inception, with only two cases that were more optimistic than Consensus. Again, at first review, the size of the errors was much reduced for IMF forecasts, but not so for Consensus. Once more, the perception that IMF forecasts are particularly erroneous does not appear to be justified by these data.

52. We now consider whether "big" programs, identified as those with more than two billion dollars in disbursement and more prominent in the media, may be the ones exhibiting below-par performance. Results in Table 7 correspond to those in Table 6 with a

[34] Loungani and others (2011) found evidence of information rigidity for that group of forecasters.

different partition of the data: big programs (above 2 billion SDRs) and others.[35] In this case, we present only the heat map while the full regression results are reproduced in Annex 4.

Figure 2. IMF and Consensus forecast errors



Source: IMF MONA and Consensus forecasts.

53. We find that the pattern for large disbursement programs is similar to that for exceptional access programs: in both cases there is some evidence of an optimistic bias in IMF forecasts of real GDP growth, slightly more convincing evidence of an optimistic bias in projections of CPI inflation, and a pessimistic bias in projections of government balance[36] and current account.

[35] The number of observations is slightly smaller than for exceptional access programs, at 188 including 20 for "big" programs.

[36] It may be that the forecasts for government balance for big programs tend to be optimistic, which would imply that their targets are more ambitious than those of small programs.

Table 7. Regression results corresponding to equation (4)

Variables	Current year forecast				One-year-ahead forecast			
	RGDP	PCPI	GGB	BCA	RGDP	PCPI	GGB	BCA
Coefficient								
β_0	0.517	1.699	1.665	2.340	-0.851	-0.885	5.479	1.917
$\beta_0 + \beta_{Big}$	-2.208	-1.377	3.541	1.002	-1.139	0.471	1.475	-0.318
$\beta_0 + \beta_{rev}$	0.984	0.905	-0.168	0.836	-0.320	0.417	4.791	0.900
$\beta_0 + \beta_{rev} + \beta_{Big} + \beta_{rev\text{-}Big}$	0.217	-0.640	2.512	1.150	-0.636	0.273	0.450	-0.983
First Review Impact								
β_{rev}	0.467	-0.794	-1.833	-1.504	0.531	1.302	-0.688	-1.017
$B_{rev} + \beta_{rev\text{-}Big}$	2.426	0.737	-1.030	0.147	0.503	-0.198	-1.024	-0.664

Pessimistic p<0.01
Pessimistic 0.01<p<0.1
Pessimistic non significant
Optimistic non significant
Optimistic 0.01<p<0.1
Optimistic p<0.01

54. We also check (Table 8) whether the type of arrangement plays a role in the forecast error bias (as suggested by Baqir and others, 2005).[37] General Resources Account-type programs (EFF, PCL, PLL, and SBA) show a weak optimistic bias in forecasts of GDP. This is not really surprising as most of the exceptional access programs (and big programs as well) are stand-by arrangements in the GRA group. Poverty Reduction and Growth Trust programs appear more prone to an optimistic bias in CPI inflation forecasts. Once more, the pessimistic bias is confirmed for government balance and current account balance. And the corrective role of program reviews is again quite apparent.

Table 8. Regression results corresponding to equation (5)

Variables	Current year forecast				One-year-ahead forecast			
	RGDP	PCPI	GGB	BCA	RGDP	PCPI	GGB	BCA
Coefficient								
β_0	0.716	1.970	0.868	1.466	-0.426	1.011	6.545	1.265
$\beta_0 + \beta_{GRA}$	-0.154	0.686	2.906	2.976	-1.149	-3.405	1.307	0.618
$\beta_0 + \beta_{rev}$	1.142	0.843	-1.876	0.689	-0.627	0.720	5.496	1.043
$\beta_0 + \beta_{rev} + \beta_{GRA} + \beta_{rev\text{-}GRA}$	0.839	0.687	2.709	1.582	0.333	-0.417	1.080	-0.944
First Review Impact								
β_{rev}	0.426	-1.127	-2.744	-0.778	-0.201	-0.291	-1.049	-0.222
$B_{rev} + \beta_{rev\text{-}GRA}$	0.992	0.000538	-0.197	-1.394	1.482	2.988	-0.227	-1.562

Pessimistic p<0.01
Pessimistic 0.01<p<0.1
Pessimistic non significant
Optimistic non significant
Optimistic 0.01<p<0.1
Optimistic p<0.01

Note: As in equation (5), GRA denotes the following arrangements: EFF, PCL, PLL, and SBA, and PRGT denotes ECF, ESAF, ESF, PRGF, PSI, and SCF.

[37] Full regression results are presented in Annex 5.

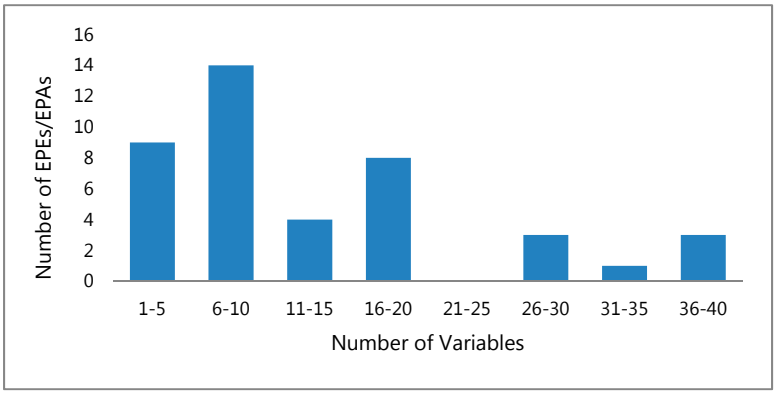
V. Survey of EPAs' and EPEs' Assessments of Program Projections

55. In studies and guidance notes issued by PDR (and SPR more recently) the IMF has recognized the value in assessing the quality of projections made in the context of IMF programs, even though the emphasis is given to projections in the context of either longer-term program engagements or exceptional access arrangements. According to the Revised Operational Guidance Note on Ex Post Assessments of Members with a Longer-Term Program Engagement (IMF, 2006b), the assessments should address "How accurate were program projections of key assumptions and objectives, and were the risks correctly identified?"[38] The most recent update of the same guidance note (IMF, 2010a) retains the same text.[39] Similarly, the guidance note for Ex Post Evaluations of Exceptional Access Arrangements (IMF, 2005) and its revised version (IMF, 2010b) stress that the EPE "should examine the macroeconomic outlook and the financing projections under the original program and compare them with the corresponding out-turns."[40]

56. Based on analysis of the 42 EPAs and EPEs that were completed between 2006 and 2013, we assess whether such guidance was followed and to what extent. For this purpose we created a template to summarize our findings of the textual analysis of the EPAs and EPEs in the sample. The template is presented in Annex 6.

57. The set of variables considered in EPEs and EPAs varies considerably in our sample: it goes from 2 (in the case of Argentina 2006, where only GDP and inflation projections are studied) up to 40, with a larger proportion of documents focusing on a smaller number of variables and an average of about 13 (Figure 3).

Figure 3. Distribution of number of variables considered



Source: IMF EPAs and EPEs, 2006–2013.

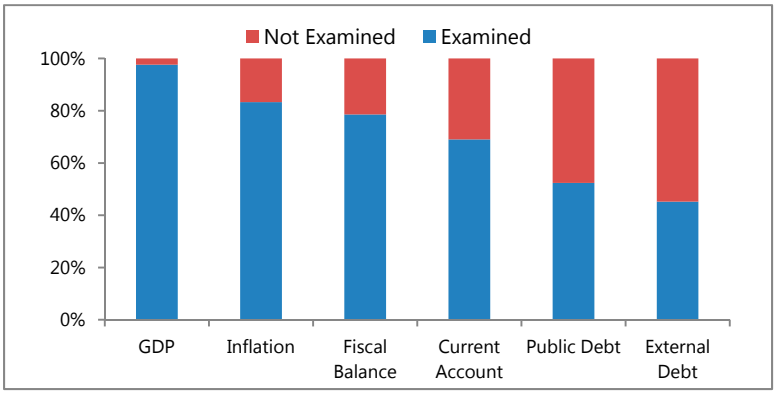
[38] IMF (2006b), p. 3.

[39] IMF (2010a), p. 3.

[40] IMF (2005), p. 3; IMF (2010b), p. 4.

58. The main variables included in the analysis are GDP, inflation, fiscal balance, external current account, public debt, and external debt. The accuracy of GDP growth projections is studied in almost all the cases in the sample; of inflation and fiscal balance projections in about 80 percent; and of external debt projections in about 50 percent (Figure 4).

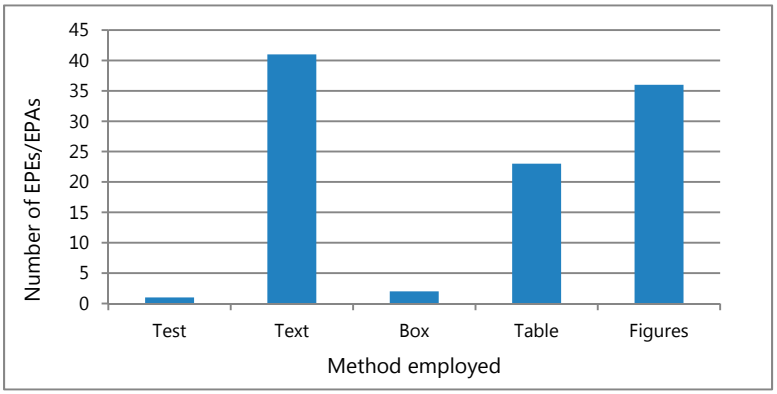
Figure 4. Main variables examined in EPEs and EPAs



Source: IMF EPAs and EPEs, 2006–2013.

59. How forecast accuracy is studied in these documents varies significantly. All of them touch on the issue: 41 documents devote some sentences in the main text, 2 have a text box, 23 present a table, and 36 show 1 or more figures, but in only one case are statistical tests employed and results presented (Figure 5). In most cases, the treatment of forecast performance is merely a list of unexpected shocks that justify the deviation from the original projection. Since these studies do not attempt to identify any systematic error, they have little to offer as learning exercises on this topic.

Figure 5. Analysis of accuracy in EPEs and EPAs

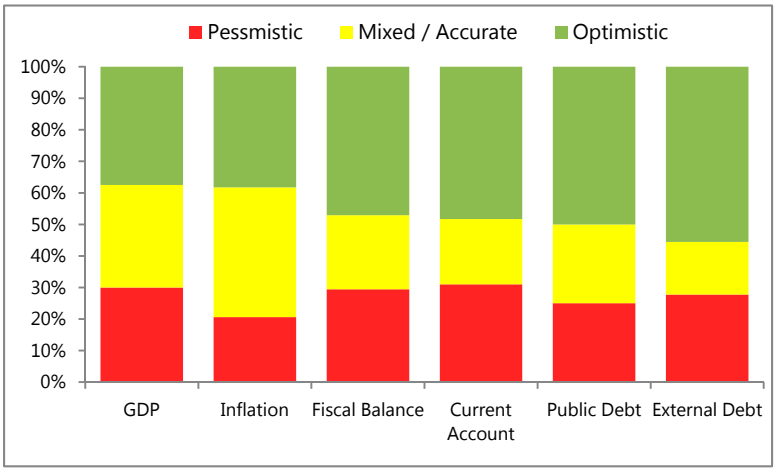


Source: IMF EPAs and EPEs, 2006–2013.

60. The results obtained in the accuracy analysis are synthesized in Figures 6, 7, and 8. Since, as mentioned above, the documents provide no formal statistical test of forecast bias (except in one instance), we look at the projection history of each program. We classify each set of short-term projections (on the six main variables listed above, where available)

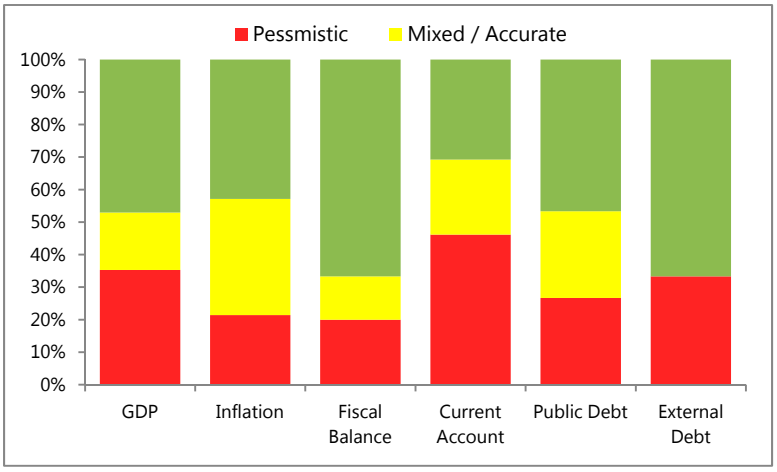
produced during the program period (inception and various reviews) as "optimistic, pessimistic, or mixed" according to a qualified majority rule. If the majority of projections turn out to be optimistic, then the overall set is classified as "optimistic,"[41] and if the errors are predominantly on the pessimistic side, then the set of projections is classified as "pessimistic." The classification takes into consideration the sample size: for example, with a sample size of three (e.g., inception plus two reviews), if the first two projections are optimistic while the last is pessimistic, the set is classified as "mixed."

Figure 6. Distribution of biases in EPEs and EPAs



Source: IMF EPAs and EPEs, 2006–2013.

Figure 7. Distribution of biases in EPEs



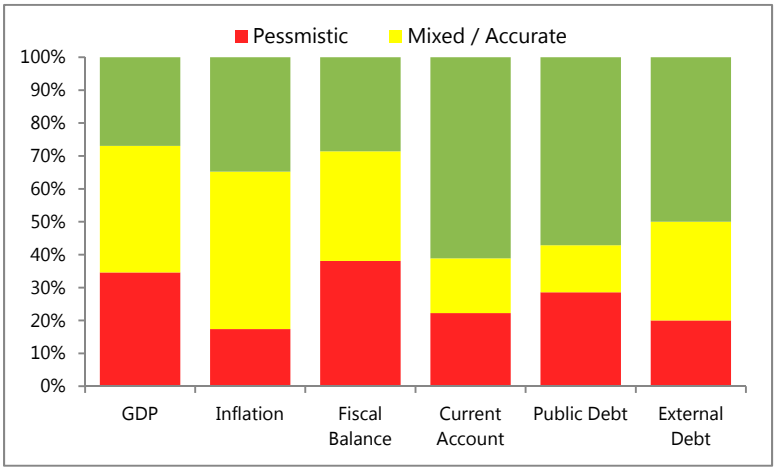
Source: IMF EPAs and EPEs, 2006–2013.

[41] The definition of "optimistic" changes depending on the variable considered: GDP, forecast > out-turn; inflation, forecast < out-turn; government balance, forecast > out-turn (typically an optimistic forecast implies a smaller deficit than the out-turn); external current account, forecast > out-turn (again, smaller forecast deficit than out-turn); public debt, forecast < out-turn; external debt, forecast < out-turn.

61. Though the results differ somewhat when considering EPEs and EPAs separately, they still suggest that the optimistic bias found by several formal statistical studies is not uniformly present in all programs. An optimistic bias is more evident in GDP and fiscal balance projections in the context of exceptional access arrangements (as recorded in EPEs), while an optimistic bias in inflation forecasts would seem quite strong in longer-term program engagements (as summarized in EPAs). This is consistent with the formal econometric results reported in Section IV above.

62. The causes of the errors in forecasting as reported by EPEs and EPAs can be grouped according to the main factors identified in the literature: (i) poor data; (ii) slippages in program implementation; and (iii) exogenous shocks. In addition, in several instances the documents refer, explicitly or obliquely, to overly optimistic or overly pessimistic (conservative) forecasts. Poor data as a cause of errors in forecasts includes factors like large revisions, poor timeliness of data release, methodological shortcomings (on the part of the authorities), and deliberate misrepresentation. These factors account for about 10 percent of the justifications provided for the forecasting errors. Slippages in implementation (including administrative limitations) are mentioned in about 25 percent of the cases. Exogenous shocks (weather, commodity prices, Russian crisis, donor assistance failure, etc.) are the most frequent "explanation," mentioned in more than 40 percent of the cases, while excessive optimism (15 percent) and pessimism (5 percent) close the list of reasons for the observed errors.

Figure 8. Distribution of biases in EPAs



Source: IMF EPAs and EPEs, 2006–2013.

63. According to the EPE and EPA guidelines, the final document must include comments expressed by the country authorities on the analysis contained in the EPE or EPA. Out of 42 documents, 32 include an annex with the authorities' comments. The study of these annexes is interesting. Only 7 of them touch upon program forecasts and 6 of the 7 are quite critical of the interpretation contained in the document. In 4 of them, the authorities complain that the projections of GDP growth and/or fiscal revenues were unjustifiably overly optimistic, stemming from a poor understanding of the economy, and—worse—that unduly

strict fiscal targets slowed down the recovery by depriving the government of needed fiscal space. The other 2 annexes with critical comments on the Fund's interpretation complain of the opposite: they say that forecasts were overly pessimistic and that recovery was much faster than projected.

64. All in all, the issue of forecast accuracy is explicitly addressed by the authorities in 15 percent of the sample of EPEs/EPAs. Such a proportion may underestimate the actual relevance of the issue for country authorities. Indeed, the annexes are devoted to comments on the report, which is already an assessment (more precisely a self-assessment) of the IMF program. In some cases where the document already criticizes the accuracy of forecasts stressing the implications of the errors, authorities may have decided not to add their comments as they agreed with the analysis.

VI. CONCLUSIONS

65. The wide perception of an optimistic bias in IMF forecasts is only partly confirmed for countries with IMF programs approved between 2002 and 2011. Across all these programs on average, forecasts of real GDP growth showed only a non-significant optimistic bias. By contrast, for the countries with exceptional access arrangements, the forecast bias at program inception was optimistic and significant. This fact may explain the generalized public perception of the optimistic bias: exceptional access arrangements have, over the years, received considerable media attention.

66. Inflation forecasts suffered from optimism more consistently, especially in cases of concessional lending arrangements like the PRGF.

67. Persistent biases emerge quite clearly in the forecasts of government balance and current account balance. There is strong evidence that over the evaluation period these forecasts suffered from persistent pessimism—i.e., larger deficit or smaller surplus than actual out-turns. The results suggest, on average, that fiscal and international reserves targets were set so as to be more easily met.

68. The data show that once a program was in place, much of the optimistic bias in real GDP growth and CPI inflation forecasts was corrected during the first review. This is consistent with the view that the incentives to present an optimistic scenario are no longer a factor in the discussions with the country's authorities.

69. We also determined that the Fund's GDP growth and inflation forecasts did not encompass all the information available at the time they were prepared. In particular, econometric tests suggest that IMF program forecast errors could be reduced by carefully scrutinizing Consensus forecasts (where they exist) with a view to identifying the driving forces behind them.

70. EPEs and EPAs are a potentially powerful tool for self assessment. However, the analysis of forecasts contained in EPEs and EPAs appeared to be somewhat pro-forma in most cases. More rigorous analysis would help the institution learn from past experience.

REFERENCES

Artis M.J., 1988, "How accurate is the World Economic Outlook? A Post Mortem on Short-Term Forecasting at the International Monetary Fund," Staff Studies for the World Economic Outlook (Washington D.C.: International Monetary Fund, July), pp. 1–49.

Artis M.J., 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).

Aslund, A., 2013, "The IMF Shouldn't Abandon Austerity." Available at www.bloomberg.com/news/2013-04-02/imf-shouldn-t-abandon-austerity.html

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

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

Coibion, Olivier and Yurij Gorodnichenko, 2012, "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," IMF Working Paper WP/12/296, December (Washington: International Monetary Fund).

De Resende, Carlos, 2014, "An Assessment of IMF Medium-Term Forecasts of GDP Growth, IEO Background Paper No. BP/14/01 (Washington: International Monetary Fund).

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

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

Goldstein, Morris, 1986, "The Global Effects of Fund-Supported Adjustment Programs," IMF Occasional Paper, 42, March (Washington: International Monetary Fund)

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

International Monetary Fund, 2005, "Ex Post Evaluation of Exceptinal Access Arrangements-Guidance Note," SM/05/303 (Washington: International Monetary Fund).

International Monetary Fund, 2006a, "Argentina-Ex Post Assesment of Longer-term Engagement and Ex Post Evaluation of Exceptinal Access," SM/06/247 (Washington: International Monetary Fund).

International Monetary Fund, 2006b, "Ex Post Assessments of Members with a Longer-Term Program Engagement-Revised Operational Guidance Note," SM/06/395 (Washington: International Monetary Fund).

International Monetary Fund, 2010a, "Ex Post Assessments of Members with a Longer-Term Program Engagement-Revised Guidance Note," SM/10/45 (Washington: International Monetary Fund).

International Monetary Fund, 2010b, "Ex Post Evaluation of Exceptinal Access Arrangements-Revised Guidance Note," SM/10/44 (Washington: International Monetary Fund).

International Monetary Fund, 2011, "2011 Triennial Surveillance Review-External Commentary-A Short Note on Surveillance and How REforms in Surveillance Can Help IMF to Promote Global Financial Stability," SM/11/237 (Washington: International Monetary Fund).

Loungani, Prakash, Herman Stekler, and Natalia Tamirisa, 2011, "Information Rigidities in Growth Forecasts: some cross-country evidence," IMF Working Paper WP/11/125, December (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.

Rosnick, David and Mark Weisbrot, 2007, "Political Forecasting? The IMF's Flawed Growth Projections for Argentina and Venezuela," April (Washington: Center for Economic and Policy Research). Available at www.cepr.net/index.php/publications

Sterne, Gabriel, 2013a, "Time for a transparency revolution at the IMF." Available at http://ftalphaville.ft.com/2013/04/19/1466472/guest-post-time-for-a-transparency-revolution-at-the-imf/.

Sterne, Gabriel, 2013b, "Regrets, IMF had a few." Available at (http://infoviewer.info/infodisplay/story/ieJnEQlDsJT4eZCOHNiAQ13.html?CU=imf5992&APP=6)

Timmermann, A., 2006, "An Evaluation of the World Economic Outlook Forecasts," IMF Working Paper No. 06/59, March (Washington: International Monetary Fund).

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).

ANNEX 1. IMF PROGRAMS AND PROGRAM REVIEWS

In its mission to promote global economic stability, the International Monetary Fund carries out three primary functions: Surveillance, Financing, and Technical Assistance.[1] IMF Financing is provided to member countries in order to help them tackle balance of payments problems, stabilize their economies, and restore sustainable economic growth. A program supported by IMF financing is a set of measures and policies that the country authorities design, in cooperation with the IMF, which are aimed at achieving specific, quantified goals in support of a member's overall economic program, including the resolution of the underlying balance of payments problem. There are different facilities and instruments that the IMF uses to provide financing in support of the member's program, depending on the nature of the support that is needed and the level of development of the country.[2] Fund financing is normally granted through Fund arrangements in support of a member's program, which constitute unilateral decisions approved by the IMF's Executive Board. Financing under an arrangement (e.g., SBA) is typically disbursed to the member in a number of installments over the life of the IMF-supported program, with each installment subject to implementation of program conditionality (i.e., performance criteria, targets and benchmarks) established by the Executive Board.

Normally, an IMF team will visit the program country shortly before each Executive Board review of the member's progress in implementing the program. During these missions, the team will discuss with the member country authorities whether the program conditionality has been met and will discuss future program conditionality. Naturally, for each of these visits (typically between three and six months apart), the IMF team will produce an updated set of macroeconomic forecasts to help inform the new targets.

There are two types of reviews: a standard program review and a financing assurance review—a review of performance regarding the reduction of arrears to private creditors.

[1] For an exhaustive overview of the IMF mission and work, visit www.imf.org/external/about/ourwork.htm.

[2] A list of the currently available facilities along with a brief description is given at www.imf.org/external/about/lending.htm.

ANNEX 2. DESCRIPTION OF THE MONA DATABASE

The Monitoring of Fund Arrangements (MONA) database is a repository of qualitative and quantitative information regarding the bilateral engagements of the Fund with its member countries. The arrangements covered by the MONA are the SBA, SAF, ESAF, EFF, PRGF, ESF, ECF, SCF, SBASBCF, PCL, FCL, PLL and PSI. The database is designed to provide the Fund with a central source of information on program design and performance vis-à-vis program targets.

The database is divided into two periods: 1993–2003 and 2002–present, as a result of a reclassification and restructuring of several economic variables that occurred in the early 2000s. The database covering arrangements approved from 1993 to 2001 is available at www.imf.org/external/np/pdr/mona/HistoricalData.aspx. It contains a complete set of data related to each program and its subsequent reviews until the program expired. The database covering arrangements approved since2002, available at www.imf.org/external/np/pdr/mona/index.aspx, records—in so far as macroeconomic variables are concerned—for each program, only the data related to the most recent review. The data used in this study were obtained from the Strategy and Policy Review Department that manages the database. Non macroeconomic data is available for each review.

The following excerpt is from the Fund intranet access to MONA at www-intranet.imf.org/departments/SPR/Databases/Pages/MONA.aspx.

DATA COVERAGE

Information covered by the MONA falls into two general categories:

- Loan arrangement description—details of the loan arrangement between the IMF and the recipient country: goals and reform strategies, loan disbursement schedule, review schedules, and performance criteria/benchmarks
- Macroeconomic overview—overview of historical economic performance and prospects/forecasts partly based on expected program performance
    - Output and prices
    - Current account and balance of payments
    - Monetary and financial market data
    - General government fiscal balance and debt.

REVIEWS AND REVISION SCHEDULES

The IMF reviews are generally held twice a year but sometimes quarterly for a few countries; and depending on the approval date, the first review can occur in the same year as the approval year. Also, most reviews were scheduled on either the third or fourth week of the review month (actual completion dates vary significantly from schedule), though there seems

to be no systematic pattern on which month the reviews occur. The Fund intranet access to MONA is at www-intranet.imf.org/departments/SPR/Databases/Pages/MONA.aspx.

PAGE CONTENT

MONITORING OF FUND ARRANGEMENTS (MONA) DATABASE

The Monitoring of Fund Arrangements, or MONA database, is an IMF-maintained database established in 1993 used to track comparable data on the economic objectives and outcomes of Fund-supported arrangements. It is a comprehensive and readily accessible source of information starting from program approval by the Executive Board to the arrangement's last review[42]. The database serves three purposes:

(i) Enhances the Fund's institutional memory -- it is the only electronic database on program design, compliance, and economic targets and developments.

(ii) Improves the IMF's capacity to respond in a timely manner to questions about country experiences under Fund-supported programs -- for example the ability to provide cross-country data on specific aspects of such programs.

(iii) Provides a time-saving source of information for the ex post evaluation of program design and progress made by countries under programs.

More specifically, the MONA database contains data pertaining to programs under SBA, SAF, ESAF, EFF, PRGF, and PSI on the following topics:

- **Description*:*** An overview of the type of loan arrangement type, the duration of the arrangement, total access to Fund financing, whether the arrangement is being treated as precautionary, and provides comments related to augmentations, extensions, and other details of established programs;
- **Program Goals and Reform Strategies:** A description of the links between program goals and strategies;
- **Purchases*:*** Covers the proposed schedule for disbursements, and the revised, and actual disbursements under the arrangement;
- **Reviews:** Covers the original schedule for Executive Board reviews, the revised schedule, and the actual program review dates;
- **Quantitative Performance Criteria:** Covers the original schedule of criteria to be applied under the program, revisions, adjustments and actual values of quantitative performance criteria and the actual test dates under the program;

[42] This statement is not precise: as mentioned above—in so far as macroeconomic variables are concerned—for each program, only the data related to the most recent review are accessible. This section: "Page Content" is extracted from the IMF Intranet website: www-intranet.imf.org/departments/SPR/Databases/Pages/MONA.aspx.

- **SPC/SAC:** Covers information on structural performance criteria, or structural assessment criteria, including relevant test dates, and outcomes;
- **PA:** Covers information on structural prior actions, including relevant test dates, and outcomes;
- **SB:** Covers information on structural benchmarks, including relevant test dates, and outcomes;
- **Macroeconomic:** An overview of macroeconomic variables covering an eight-year span from "t-3" through "t+4," with "t" being the initial program year.

Guide to Classification of Structural Conditions (2009)

MONA: ECONOMIC SECTOR CLASSIFICATION OF STRUCTURAL CONDITIONS

1. General government

1.1 Revenue measures, excluding trade policy

Includes introduction or elimination of taxes and non-tax revenues, changes in rates and exemptions, rationalization of tax incentives, and revisions to the tax code. Changes in tariffs, surcharges and other levies on imports and exports; export drawback systems; export processing zones and similar tax measures that primarily aim to affect international trade should be recorded under international trade policy (category 8). Measures pertaining to VAT on imports, however, should be recorded under revenue measures.

1.2 Revenue administration, including customs

All measures aimed at the operations of the tax and non-tax revenue administrations are covered under this category. It includes, in particular, measures aimed at improving the effectiveness, efficiency, and governance of the bodies charged with revenue administration and tax courts. Short-term measures aimed at increasing revenue performance (such as a temporary increase in tax collection efforts) should be recorded under revenue measures (1.1).

1.3 Expenditure measures, including arrears clearance

All measures aimed at capping or changing the level and composition of government expenditure, including the preparation of medium-term expenditure frameworks, policies aimed at reaching certain levels of social spending, and legal provisions aimed at capping expenditures (such as a "fiscal responsibility law" with specific expenditure ceilings). This category also includes the elimination (through payment or restructuring) of domestic and external payments arrears as well as the clearance of public enterprise arrears (including cross debts).

1.4 Combined expenditure and revenue measures

In some cases, program conditionality is phrased in such a way that it covers (or could cover) both expenditures and revenues (for instance, deficit targets). These should be recorded in category 1.4.

1.5 Debt management

This category covers all policies related to the size, composition, and terms of domestic and external public sector debt (with the exception of arrears clearance, which should be recorded under 1.3), including policies related to debt guarantees issued by the government; public enterprise debt; and contingent fiscal liabilities. Measures aimed at the organization, efficiency, effectiveness, and governance of bodies charged with public debt policy are also included.

1.6 Expenditure auditing, accounting, and financial controls

This category covers reforms of public financial management systems, Treasury reforms, audits, and other financial control systems for government and public sector bodies (excluding public enterprises, which would be recorded under 5.2). It includes, for instance, the introduction of new expenditure control mechanisms (such as computerized expenditure authorization systems), and reforms of government institutions aimed at strengthening spending controls.

1.7 Fiscal transparency

All measures aimed at creating or improving fiscal transparency for the public and parliament should be recorded in this category. They could include the publication of government or public enterprise accounts and audit reports; public procurement transactions; and procedures and bids received in the context of a privatization.

1.8 Budget preparation

All actions related to the preparation of the government budget (including submission to, and approval by the cabinet or parliament) should be recorded here. This category also covers measures such as the introduction of a new organic budget law, revisions to the budget nomenclature and changes in procedures related to budget preparation.

1.9 Inter-governmental relations

This category includes measures such as fiscal decentralization and agreements between central and local governments on spending limits and tax authority. Understandings between the central government and the central bank on monetary policy or prudential supervision should be recorded under 2 or 6, respectively.

2. Central bank

2.1 Central bank operations and reforms

This category covers monetary policy, money market operations, exchange rate policy, reserve policy, and organizational reforms of the central bank (including, for instance, adoption of a new central bank law). Measures in the area of debt management, prudential supervision, and exchange controls, which are often implemented by the central bank, should be recorded under 1.5, 6, and 7, respectively.

2.2 Central bank auditing, transparency, and financial controls

Including control procedures for reserve management and investment rules.

3. Civil service and public employment reforms, and wages

This category covers a broad range of measures related to the public administration, including, e.g. changes in the numbers of civil servants or the composition of the civil service; reforms of pay scales and promotion systems; payroll reforms and the elimination of ghost workers; the establishment, abolition or restructuring of ministries and other public bodies (including local governments and semi-public bodies such as government-funded institutions, but excluding public enterprises and the central bank); and the introduction of public wage caps.

4. Pension and other social sector reforms

4.1 Pension reforms

This includes measures relating to eligibility, employer and employee contributions, and actuarial reserves of public or private pension funds. Measures related to prudential supervision of pension funds are recorded under 6.

4.2 Other social sector reforms

Actions or measures related social safety nets or access to education and health (including insurance) are recorded under this heading. Measures related to social spending levels, however, are recorded under 1.3, and policies related to numbers of public workers in the social sectors are normally recorded under 3.

5. Public enterprise reform and pricing (non financial sector)

5.1 Public enterprise pricing and subsidies

Including limits on state aid to public enterprises, the adoption of new energy prices and pricing formulas for public utilities, and similar measures.

5.2 Privatization, public enterprise reform and restructuring, other than pricing

This includes all steps leading to privatization and public enterprise reforms (for instance, a due diligence in preparation for a public offer for sale). Public enterprise reforms could include measures such as commercialization of operations; changes in management and staff; and financial audits.

6. Financial sector

6.1 Financial sector legal reforms, regulation, and supervision

This category covers all measures related to prudential supervision and financial market regulation. Restructuring (organizational and/or financial) and privatization of financial institutions, which may constitute a step in the implementation of financial regulation, are recorded under 6.2.

6.2 Restructuring and privatization of financial institutions

This includes the closure of financial institutions for prudential reasons.

7. Exchange systems and restrictions (current and capital)

This includes changes in the exchange regime, payments restrictions on current and capital account transactions (including, e.g., margin requirements on letters of credit), and reforms of the foreign exchange market (including, for instance, introduction of screen-based interbank foreign exchange transactions). Exchange rate policy (other than changes in the exchange regime) is recorded under 2.1.

8. International trade policy, excluding customs reforms

All tariff and non-tariff measures, tariff exemptions, and export processing zones are covered. This category does not include measures that improve the operations of customs administration unless they change either tariff rates or the revenue base on which tariffs are collected.

9. Labor markets, excluding public sector employment

This category covers wage policies (including caps on minimum wages), labor market regulations (including hiring and firing rules, redundancy compensation, dispute settlement), work time flexibility (including rules on fixed-term contracts).

10. Economic statistics (excluding fiscal and central bank transparency)

This covers the generation or improvement of published data (e.g., national accounts, balance of payments statistics) and data produced to enable the authorities and the Fund team to monitor implementation of the program.

11. Other structural measures

11.1 Private sector legal and regulatory environment reform (non financial sector)

Under this category fall measures such as reform of the judiciary (excluding tax courts, which would be recorded under 1.2); investment promotion policies (except those related to export processing zones, which are recorded under 8); reforms of ownership rights and bankruptcy law, and similar measures that directly affect the investment climate. Also included are measures related to private sector pricing.

11.2 Natural resource and agricultural policies (excluding public enterprises and pricing)

All measures related to the development, organization and regulation of natural resource extraction and the agricultural sector. Examples include: adoption of a standard mining contract; finalization of a national energy strategy; liberalization of marketing arrangements in the agricultural sector; the adoption of a system for minimum producer prices in the agricultural sector; the adoption of a national fertilizer distribution strategy.

11.3 PRSP development and implementation

Including measures such as the completion of a household poverty survey, a country poverty assessment, or publication of an anti-poverty strategy.

11.4 Anti-corruption legislation/policy

Includes the preparation of an anti-corruption strategy, the establishment of an anti-corruption commission, the adoption of financial disclosure rules, the introduction or strengthening of public bodies investigating corruption cases.

ANNEX 3. PROGRAMS COVERED BY THE ECONOMETRIC ANALYSIS IN SECTION IV

Country	Arrangement Class	Year
Albania	PRGT	2002
Albania	PRGT	2006
Angola	GRA	2009
Antigua and Barbuda	GRA	2010
Argentina*†	GRA	2000
Argentina*†	GRA	2003
Armenia	PRGT	2005
Armenia*	GRA	2009
Armenia	PRGT	2010
Belarus*†	GRA	2009
Benin	PRGT	2005
Benin	PRGT	2010
Bolivia	GRA	2003
Bosnia and Herzegovina	GRA	2002
Bosnia and Herzegovina	GRA	2009
Brazil*†	GRA	2002
Bulgaria	GRA	2004
Burkina Faso	PRGT	2003
Burkina Faso	PRGT	2007
Burkina Faso	PRGT	2010
Burundi	PRGT	2004
Burundi	PRGT	2008
Cameroon	PRGT	2005
Cape Verde	PRGT	2002
Cape Verde	PRGT	2006
Cape Verde	PRGT	2010
Central African Republic	PRGT	2006
Colombia	GRA	2003
Colombia	GRA	2005
Comoros	PRGT	2009
Costa Rica*	GRA	2009
Côte d'Ivoire	PRGT	2009
Croatia	GRA	2003
Croatia	GRA	2004
Djibouti	PRGT	2008
Dominica	GRA	2002
Dominica	PRGT	2003
Dominican Republic	GRA	2003
Dominican Republic	GRA	2009
Ecuador	GRA	2003
El Salvador	GRA	2010
Ethiopia	PRGT	2009
FYR Macedonia	GRA	2005
Gabon	GRA	2004
Gabon	GRA	2007
Georgia	PRGT	2004
Ghana	PRGT	2003
Greece*†	GRA	2010
Grenada	PRGT	2006
Grenada	PRGT	2010
Guatemala	GRA	2002
Guinea	PRGT	2007
Guinea-Bissau	PRGT	2010
Honduras	PRGT	2004
Honduras	GRA	2010
Hungary*†	GRA	2008
Iceland*	GRA	2008
Iraq	GRA	2007
Iraq†	GRA	2010
Jordan	GRA	2002
Kyrgyz Republic	PRGT	2005
Latvia*	GRA	2008
Lesotho	PRGT	2010
Liberia*	PRGT	2008
Madagascar	PRGT	2006
Malawi	PRGT	2005
Malawi	PRGT	2010
Maldives	GRA	2009
Mali	PRGT	2004
Mali	PRGT	2008
Mauritania	PRGT	2006
Mauritania	PRGT	2010
Moldova	PRGT	2006
Moldova	PRGT	2010
Mongolia*	GRA	2009
Mozambique	PRGT	2004
Mozambique	PRGT	2007
Mozambique	PRGT	2010
Nicaragua	PRGT	2007
Niger	PRGT	2005
Niger	PRGT	2008
Nigeria	PRGT	2005
Paraguay	GRA	2003
Paraguay	GRA	2006
Peru	GRA	2004
Peru	GRA	2007
Republic of Congo	PRGT	2004
Republic of Congo	PRGT	2008
Romania	GRA	2004
Rwanda	PRGT	2002
Rwanda	PRGT	2006
Rwanda	PRGT	2010
São Tomé and Príncipe	PRGT	2005
São Tomé and Príncipe	PRGT	2009
Senegal	PRGT	2003
Senegal	PRGT	2007
Senegal	PRGT	2010
Serbia*†	GRA	2009
Seychelles	GRA	2008
Seychelles	GRA	2009
Sierra Leone	PRGT	2001
Sierra Leone	PRGT	2006
Sierra Leone	PRGT	2010
Solomon Islands	PRGT	2010
Tajikistan	PRGT	2002
Tanzania	PRGT	2007
The Gambia	PRGT	2007
Togo	PRGT	2008
Uganda	PRGT	2006
Ukraine*†	GRA	2008
Ukraine*†	GRA	2010
Uruguay	GRA	2002
Uruguay*	GRA	2005
Zambia	PRGT	2004

Note: * Arrangements classified as Exceptional Access; † Arrangements classified as Big Programs (disbursements over 2 billion SDRs).

ANNEX 4. REGRESSION RESULTS CORRESPONDING TO TABLE 7 IN THE MAIN TEXT

	Current year forecast				One year ahead forecast			
VARIABLES	RGDP	PCPI	GGB	BCA	RGDP	PCPI	GGB	BCA
$\beta_{till\text{-}hor}$	-0.0236	-0.00941	-0.0462	-0.0185	0.0208	0.0456	-0.0370	-0.0128
	(0.0181)	(0.0166)	(0.0445)	(0.0362)	(0.0249)	(0.0357)	(0.0270)	(0.0371)
B_{rev}	0.467	-0.794**	-1.833	-1.504	0.531	1.302	-0.688	-1.017
	(0.430)	(0.338)	(1.254)	(1.133)	(0.550)	(1.358)	(0.431)	(1.349)
β_{Big}	-2.725*	-3.076**	1.876	-1.338	-0.288	1.356	-4.004	-2.236
	(1.472)	(1.352)	(2.152)	(2.256)	(1.422)	(2.049)	(2.871)	(2.956)
$\beta_{rev\text{-}Big}$	1.959***	1.531*	0.803	1.651	-0.0282	-1.500	-0.337	0.353
	(0.744)	(0.930)	(1.756)	(1.323)	(0.757)	(1.416)	(1.002)	(1.631)
β_0	0.517	1.699**	1.665*	2.340	-0.851	-0.885	5.479**	1.917
	(0.751)	(0.860)	(0.888)	(1.653)	(1.164)	(2.207)	(2.455)	(1.846)
Observations	188	188	142	96	188	188	142	96
Arrangements	94	94	71	48	94	94	71	48
Rho	0.665	0.828	0.676	0.533	0.461	0.196	0.978	0.736
Wald Prob > chi2	0.000251	0.0760	0.203	0.749	0.813	0.715	0.0831	0.611
Coefficient								
β_0	0.517	1.699	1.665	2.340	-0.851	-0.885	5.479	1.917
Standard Error	0.751	0.860	0.888	1.653	1.164	2.207	2.455	1.846
p-value	0.491	0.0483	0.0608	0.157	0.465	0.688	0.0256	0.299
$\beta_0 + \beta_{Big}$	-2.208	-1.377	3.541	1.002	-1.139	0.471	1.475	-0.318
Standard Error	1.455	1.315	2.016	2.102	1.446	1.930	2.037	2.582
p-value	0.129	0.295	0.0790	0.633	0.431	0.807	0.469	0.902
$\beta_0 + \beta_{rev}$	0.984	0.905	-0.168	0.836	-0.320	0.417	4.791	0.900
Standard Error	0.521	0.671	1.646	1.072	0.678	1.006	2.482	1.919
p-value	0.0588	0.178	0.919	0.435	0.637	0.679	0.0536	0.639
$\beta_0 + \beta_{rev} + \beta_{Big} + \beta_{rev\text{-}Big}$	0.217	-0.640	2.512	1.150	-0.636	0.273	0.450	-0.983
Standard Error	1.165	0.669	1.438	1.636	1.415	1.641	1.507	2.025
p-value	0.852	0.339	0.0808	0.482	0.653	0.868	0.765	0.627
First Review Impact								
β_{rev}	0.467	-0.794	-1.833	-1.504	0.531	1.302	-0.688	-1.017
Standard Error	0.430	0.338	1.254	1.133	0.550	1.358	0.431	1.349
p-value	0.278	0.0187	0.144	0.184	0.334	0.338	0.110	0.451
$B_{rev} + \beta_{rev\text{-}Big}$	2.426	0.737	-1.030	0.147	0.503	-0.198	-1.024	-0.664
Standard Error	0.621	0.869	1.058	0.858	0.611	0.749	0.912	0.941
p-value	9.46e-05	0.396	0.330	0.864	0.411	0.791	0.261	0.480

Robust standard errors in parentheses

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

Pessimistic p<0.01

Pessimistic 0.01<p<0.1

Pessimistic non significant

Optimistic non significant

Optimistic 0.01<p<0.1

Optimistic p<0.01

ANNEX 5. REGRESSION RESULTS CORRESPONDING TO TABLE 8 IN THE MAIN TEXT

	Current year forecast				One year ahead forecast			
VARIABLES	RGDP	PCPI	GGB	BCA	RGDP	PCPI	GGB	BCA
$\beta_{till\text{-}hor}$	-0.0270	-0.0105	-0.0407	-0.0221	0.0167	0.0520	-0.0262	-0.00323
	(0.0173)	(0.0153)	(0.0404)	(0.0360)	(0.0235)	(0.0366)	(0.0238)	(0.0381)
B_{rev}	0.426	-1.127***	-2.744	-0.778	-0.201	-0.291	-1.049*	-0.222
	(0.307)	(0.371)	(1.712)	(0.977)	(0.173)	(0.270)	(0.620)	(1.164)
β_{GRA}	-0.870	-1.283	2.038	1.510	-0.723	-4.416	-5.238	-0.647
	(1.153)	(1.355)	(2.074)	(2.455)	(1.429)	(3.119)	(3.254)	(2.819)
$\beta_{rev\text{-}GRA}$	0.566	1.128*	2.547	-0.617	1.683	3.279	0.822	-1.340
	(0.870)	(0.617)	(1.735)	(2.187)	(1.210)	(3.012)	(0.713)	(2.506)
β_0	0.716	1.970***	0.868	1.466	-0.426	1.011	6.545**	1.265
	(0.560)	(0.675)	(1.296)	(1.364)	(0.608)	(0.904)	(3.332)	(1.918)
Observations	206	206	158	106	206	206	158	106
Arrangements	103	103	79	53	103	103	79	53
Rho	0.665	0.824	0.680	0.530	0.478	0.195	0.976	0.741
Wald Prob > chi2	0.0504	0.0545	0.188	0.769	0.178	0.491	0.111	0.958
Coefficient								
β_0	0.716	1.970	0.868	1.466	-0.426	1.011	6.545	1.265
Standard Error	0.560	0.675	1.296	1.364	0.608	0.904	3.332	1.918
p-value	0.201	0.00352	0.503	0.282	0.484	0.264	0.0495	0.510
$\beta_0 + \beta_{GRA}$	-0.154	0.686	2.906	2.976	-1.149	-3.405	1.307	0.618
Standard Error	1.201	1.358	1.195	2.428	1.803	3.773	0.895	2.263
p-value	0.898	0.613	0.0150	0.220	0.524	0.367	0.144	0.785
$\beta_0 + \beta_{rev}$	1.142	0.843	-1.876	0.689	-0.627	0.720	5.496	1.043
Standard Error	0.503	0.523	2.074	1.106	0.637	0.998	3.400	1.826
p-value	0.0231	0.107	0.366	0.534	0.325	0.471	0.106	0.568
$\beta_0+\beta_{rev}+\beta_{GRA}+\beta_{rev\text{-}GRA}$	0.839	0.687	2.709	1.582	0.333	-0.417	1.080	-0.944
Standard Error	0.709	1.007	1.176	1.464	0.769	1.196	0.781	2.798
p-value	0.237	0.495	0.0213	0.280	0.665	0.727	0.167	0.736
First Review Impact								
β_{rev}	0.426	-1.127	-2.744	-0.778	-0.201	-0.291	-1.049	-0.222
Standard Error	0.307	0.371	1.712	0.977	0.173	0.270	0.620	1.164
p-value	0.166	0.00240	0.109	0.426	0.246	0.282	0.0905	0.849
$B_{rev} + \beta_{rev\text{-}GRA}$	0.992	0.000538	-0.197	-1.394	1.482	2.988	-0.227	-1.562
Standard Error	0.798	0.489	0.330	1.899	1.154	2.901	0.314	2.164
p-value	0.214	0.999	0.550	0.463	0.199	0.303	0.470	0.470

Robust standard errors in parentheses

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

Pessimistic p<0.01

Pessimistic 0.01<p<0.1

Pessimistic non significant

Optimistic non significant

Optimistic 0.01<p<0.1

Optimistic p<0.01

ANNEX 6. TEMPLATE FOR THE ANALYSIS OF EPAS AND EPES IN SECTION V OF THE MAIN TEXT

Type of Analysis

Country	Year	EPA/EPE?	Evaluation Variable	Analysis of Accuracy?					
				Done	Test	Text	Box	Table	Figures
A	2004	EPA	Forecast	Yes	No	Yes	No	No	Yes
B	2004	EPE	Program projections	Yes	No	Yes	No	No	Yes
C	2004	EPA	Program projections	Yes	Yes	Yes	Yes	Yes	Yes

Variables Analyzed

Country	Year	Which Variables?						
			GDP		Inflation		Current Account	
		Total #	Examined?	Bias/Error?	Examined?	Bias/Error?	Examined?	Bias/Error?
A	2004	2	Yes	Pessimistic	Yes	Mixed	No	N/A
B	2004	11	Yes	Mixed	Yes	Optimistic	Yes	Pessimistic
C	2004	9	Yes	Pessimistic	Yes	Mixed	Yes	Optimistic

Country	Year	Which Variables?						
			Public Debt		External Debt		Fiscal Balance	
		Total #	Examined?	Bias/Error?	Examined?	Bias/Error?	Examined?	Bias/Error?
A	2004	2	No	N/A	No	N/A	No	N/A
B	2004	11	Yes	Mixed	Yes	Pessimistic	Yes	Optimistic
C	2004	9	No	N/A	No	N/A	Yes	Pessimistic

Sources of Errors and Authorities' Comments

Country	Year	Descriptions of Errors		Authorities' Comments		
		Explanation	Reasons	Mention forecasts?	Rebutted or Agreed?	Reasons for Rebuttal
A	2004	Yes	Slowdown of global economy (p. x)	Yes	Rebutted	IMF too optimistic on export growth (p. y)
B	2004	No	N/A	N/A	N/A	N/A
C	2004	No	N/A	N/A	N/A	N/A