IEO
Independent Evaluation Office
of the International Monetary Fund

BACKGROUND PAPER

BP/21-01/01

Cross-Country Analysis of Program Design and Growth Outcomes: 2008–19

Jun I. Kim, Jean-Marc B. Atsebi, Kwang Y. Lee, Hasan H. Toprak, and Jiakun Li

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. ieo.imf.org

 BP/21-01/01

IEO Background Paper
Independent Evaluation Office
of the International Monetary Fund

Cross-Country Analysis of Program Design and Growth Outcomes: 2008–19

Prepared by Jun I. Kim,[*] Jean-Marc B. Atsebi,[†] Kwang Y. Lee,[‡] Hasan H. Toprak,[§] and Jiakun Li[**]

June 30, 2021

[*] Advisor, Independent Evaluation Office of the IMF.

[†] Economist, European Department of the IMF (Former Summer Intern, Independent Evaluation Office of the IMF).

[‡] Former Research Officer, Independent Evaluation Office of the IMF.

[§] Former Research Assistant, Independent Evaluation Office of the IMF.

[**] Research Officer, Independent Evaluation Office of the IMF.

Contents Page

Figures

Tables

Appendices

ABBREVIATIONS

AGBD	Actual Real GDP Growth relative to Benchmark
ASCI	Average Implementation Score of Structural Conditions
ASCD	Average Depth Score of Structural Conditions
ASCG	Average Growth-orientation Score of Structural Conditions
ATE	Average Treatment Effect
CABY	Current Account Balance (as a share of GDP)
EA	Exceptional Access Arrangement
EDY	External Debt-to-GDP Ratio
FE	Forecast Error
GEI	Government Effectiveness Index
GEXPY	General Government Expenditure (net of interest; as a share of GDP)
GRA	General Resource Account
GREVY	General Government Revenue (net of interest; as a share of GDP)
IV	Instrumental Variable Regression
LIC	Low-income Country
MONA	Monitoring of Fund Arrangement
NA	Normal Access Arrangement
OLS	Ordinary Least Square
OLS-FE	Ordinary Least Square with Fixed Effects
PBY	Fiscal Primary Balance (as a share of GDP)
PDY	Public Debt-to-GDP Ratio
PRGT	Poverty Reduction and Growth Trust
PTG	Potential Real GDP Growth
PTGBD	Potential Real GDP Growth relative to Benchmark
PUBINVY	Public Investment (as a share of GDP)
REER	Real Effective Exchange Rate
ROC	Review of Program Design and Conditionality
SC	Structural Condition
SCI	Aggregate Implementation Score of SCs
SCID	Aggregate Implementation and Depth Score of SCs
SCIDG	Aggregate Implementation, Depth and Growth-orientation Score of SCs
SOCIALY	Social Spending (as a share of GDP)
TOT	Terms of Trade
TPGR	Trading Partners' Growth Rate
WEO	World Economic Outlook
YGAP	Output Gap (in percent of trend GDP)

EXECUTIVE SUMMARY

This background paper presents and discusses the results of cross-country analysis undertaken to assess program design and growth outcomes of IMF-supported programs over the period of 2008–19. As to program design, assessment focuses on fiscal policy in initial program design and program adaptation. Specifically, the analysis examines how sustainability and growth considerations were incorporated in fiscal policy in initial program design and subsequent program adaptation, as well as how realistic were program assumptions on fiscal multipliers. For growth outcomes, the analysis investigates modeling error as a source of growth optimism and assesses the growth impact of IMF-supported programs in the short and medium runs.

The results of cross-country analysis are broadly supportive of the IMF's attention to growth in program design and the positive role played by IMF-supported programs in promoting growth in both short and medium runs. Key findings are summarized as follows:

- Fiscal policy in initial program design was responsive to public debt and counter-cyclical in nature in General Resource Account (GRA) programs but less clearly so in Poverty Reduction and Growth Trust (PRGT) programs. In program adaptation, fiscal policy paid good attention to growth and sustainability in both GRA and PRGT programs with sustainability considerations being generally stronger in the former.

- Initial program assumptions on fiscal multipliers were more realistic than when updated in GRA programs while the opposite holds in PRGT programs. Revenue multipliers seem to have been updated with greater realism in both GRA and PRGT programs but not expenditure multipliers particularly in GRA programs.

- Modeling errors related to the relationship between fiscal adjustment and growth in GRA programs, especially in GRA programs other than crisis programs, have been an important source of growth optimism, accounting for about 30 percent of sample variation in growth forecast errors.

- IMF-supported programs have fared relatively well when assessed based on growth deviations from a benchmark that corrects for the influence of external factors. GRA (PRGT) programs were populated dominantly with negative (positive) growth deviations.

- IMF-supported programs have helped boost growth in the short run, relative to a counterfactual of no Fund engagement. Average growth gains are on the order of 0.7 percentage points and 1.1 percentage points for completed GRA and PRGT programs, respectively. Moreover, growth gains are generally larger the higher is the depth and growth orientation of structural conditions.

- Both stabilizations and reforms implemented during the program have affected post-program potential growth positively and statistically significantly. Growth-friendly fiscal adjustment involving increased public investment and social spending seems to yield lasting growth benefits. The growth impact of structural conditions has tended to be larger the greater is their depth and growth-orientation.

I. INTRODUCTION

1. This background paper presents and discusses the results of cross-country analysis undertaken for the IEO's evaluation of growth and adjustment in IMF-supported programs. The cross-country analysis draws on the methodology and findings of the *2018 Review of Program Design and Conditionality* (ROC) in many respects, but also departs in focus and scope. The sample used for cross-country analyses discussed in this paper consists of 131 IMF-supported programs across 75 countries, approved and scheduled to be completed between September of 2008 and the first quarter of 2020. The findings from cross-country analysis complement those from select country case studies undertaken for the evaluation.

2. Broadly, three sets of cross-country analysis are undertaken in the paper. The first set is intended to assess the macroeconomic frameworks envisaged in program design focusing on the following two questions: (i) how sustainability and growth considerations were incorporated in initial program design and in program adaptation, and (ii) what program assumptions were made on fiscal multipliers in initial program design and how they were updated subsequently. The second set aims to evaluate program growth outcomes and assess the realism of the macroeconomic framework in program design. Related questions include: (i) how program growth outcomes fared relative to a common benchmark based on external factors alone, (ii) whether the behavioral relationship between adjustment and growth outcomes differed from those envisaged in program design, and (iii) whether modeling errors in the macroeconomic framework contributed to growth optimism. The third set of analysis is about the growth impact of IMF-supported programs in the short and medium runs. The short-run growth impact (i.e., during the program) is assessed relative to a notional counterfactual of no Fund engagement. For the medium-run impact, the analysis focuses on assessing post-program growth benefits of stabilizations and reforms implemented during the program.

3. The rest of the paper is structured as follows. Section II summarizes the evaluation sample and data used for the empirical analysis and explains the conventions used to determine the program period—i.e., starting and ending years of the program—considered in the regression analysis. Section III presents an overview of program outcomes for growth and adjustment relative to program objectives, program projections and initial conditions. Sections IV and V assess how sustainability and growth considerations were incorporated in initial program design and program adaptation, respectively, through the lens of programmed fiscal policy. Section VI discusses empirical findings on the macroeconomic framework envisaged in program design and adaptation, with a focus on program assumptions on fiscal multipliers.

4. Section VII compares program growth outcomes relative to a benchmark estimated based on exogenous external factors alone. Section VIII presents the results of cross-country analysis of adjustment and growth outcomes and assesses the realism of the macroeconomic framework in program design by comparing the results for program outcomes with those for program projections in Section VI. Section IX investigates the role and importance of macroeconomic modeling errors as a source of optimism bias in program growth projections.

Sections X analyzes the short-run growth impact of IMF-supported programs relative to a counterfactual of no Fund engagement, using a new technique to correct for sample selection bias. Section XI turns to the post-program growth impact of stabilizations and reforms implemented during the program. Section XII concludes by summarizing key findings from the cross-country analysis and their implications and suggests directions for further research.

II. EVALUATION SAMPLE AND DATA

5. The evaluation sample used for cross-country analysis consists of 131 IMF-supported programs arranged for 75 countries—approved and scheduled to be completed between September 2008 and March 2020 (Appendix I).[1] Given our focus on growth and adjustment, those programs with no program conditionality are excluded from the evaluation sample. The sample includes 54 General Resource Account (GRA) programs and 77 Poverty Reduction and Growth Trust (PRGT) programs including blended programs (Table 1). Programs that went off track and were never fully completed and GRA programs for countries in the context of the global financial crisis and the euro area crisis ("crisis programs") account for 37 percent and 18 percent of the sample, respectively.[2] Program approvals are more concentrated in the period of 2008–10, peaking at 28 program approvals in 2010 (Figure 1).

Table 1. Composition of the Evaluation Sample

	Full	GRA	PRGT
Number of countries[1]	75	33	44
Fragile states	25	2	23
Small states	11	4	7
Currency Union members	20	5	15
Number of programs	131	54	77
Completed programs	82	32	50
Off-track programs[2]	27	15	13
Quickly off-track programs[2]	22	7	15
Precautionary programs	18	10	8
Exceptional access programs	26	26	0
Crisis programs[3]	23	23	0

Source: IEO calculations.

[1] Armenia and Georgia have both GRA and PRGT programs.

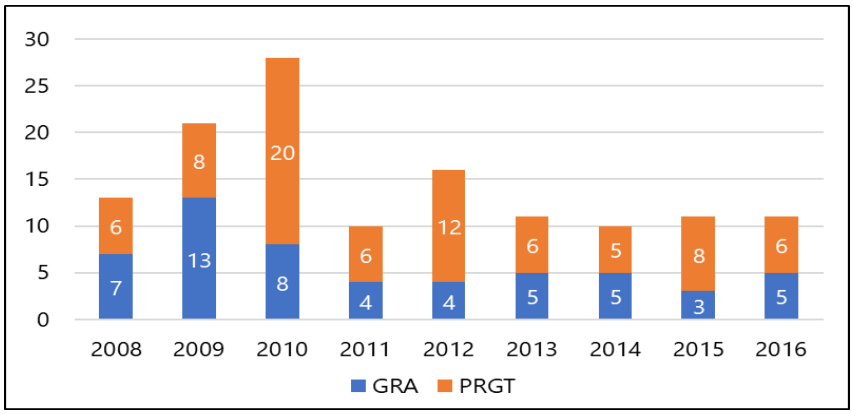
[2] Following the definition used by the 2018 ROC, "off-track programs" refer to programs where at least two reviews were completed and at least two reviews were not completed at the end of the program and "quickly off-track programs" refer to programs where at most one review was completed and at least two reviews were not completed at the end of the program.

[3] Crisis programs refer to the programs approved during 2008–09 in response to the GFC (Angola, Armenia, Belarus, Bosnia and Herzegovina, Costa Rica, Dominican Republic, El Salvador, Georgia, Guatemala, Hungary, Latvia, Mongolia, Pakistan, Romania, Serbia, Sri Lanka, and Ukraine) and five Euro Area programs arranged in response to the Euro area debt crisis (Cyprus, 2013; Greece, 2010 and 2012; Ireland, 2010; and Portugal, 2011).

[1] The sample includes programs that were subsequently cancelled or went off track and were thus never completed.

[2] The fully completed programs include the "completed and largely implemented" programs according to the classification used by the 2018 ROC (IMF, 2019b). See Table 1 for the coverage of crisis programs.

Figure 1. Distribution of Program Approvals: 2008–16



Sources: MONA database; IEO staff calculations.

6. For analytical purposes, the following conventions are used to define program duration or the length of the program in the cross-country analysis unless otherwise indicated:

- *Convention 1*. If the program is approved in the last quarter of year t, the following year t+1 is counted as the first year of the program period; otherwise year t is the first year.

- *Convention 2*. If the program is fully completed in the first quarter of year t, the previous year t-1 is considered as the last year of the program period; otherwise year t is the last year.

- *Convention 3*. For off-track programs, the last year of the program is determined based on the date of the last completed program review while applying Convention 2 above.

7. In what follows, initial or planned duration refers to the length of the program based on the ending date set at the time of program approval while actual duration refers to the length of the program based on the actual ending date. Thus, initial duration may differ from actual duration. For each program, annual averages of program projections and outcomes are calculated by using initial and actual duration, respectively. Finally, the first and last years of the program period are denoted by T and T+E, respectively, unless otherwise indicated.

8. Data used in the analysis are taken mostly from the World Economic Outlook (WEO) database and Monitoring of Fund Arrangement (MONA) database of the IMF. Actual data on economic outcomes are taken from the 2020 January vintage of the WEO database, while program projections and *real time* data are taken from various vintages depending on the specific window of time that the analysis is focusing on. In case of one-year-ahead projections, data are taken from the October vintage of the year in which projections were made. For instance, the one-year-ahead program projection for the 2010 real GDP growth rate is taken from the 2009 October vintage. Data on program approval and completion status are taken from the MONA database.

III. Growth and Adjustment in Program Design and Outcomes: An Overview

9. To set the stage for cross-country analyses in the subsequent sections, this section provides a brief overview of program objectives, initial conditions as well as growth and adjustment in program design and outcomes, with due attention to similarities and differences between GRA and PRGT programs.[3] Assessment of program objectives provides useful reference in assessing sustainability and growth considerations in program design. For growth and adjustment, cross-program averages of initial program projections (formulated at the time of program approval) are compared with those of program outcomes. Finally, the section looks at the combination of adjustment, financing and growth based on the decomposition of the BOP need, which is undertaken for select programs in the country case studies based on the same methodology used by the 2018 ROC.

10. For consistency, sample statistics discussed for comparison between program design and outcomes are constructed based on the same sample of programs. As shown in Table 1, many programs in the evaluation sample went quickly off track for which no observations are available for program outcomes given the conventions discussed in the previous section. As a result, the sample of programs used is determined by the availability of data for program outcomes and thus smaller than the entire evaluation sample. Moreover, for each program in the sample, initial program duration determined at the time of program approval may differ from actual program duration. For this reason, cross-country comparison is made largely based on annual averages over program duration. For instance, actual (programmed) fiscal adjustment in each program is calculated by dividing the actual (programmed) cumulative change in the fiscal primary during the program by actual (initial) program duration in years.

Program Objectives

11. Before discussing the results of cross-country analysis on program design and growth outcomes, it is useful to examine what program objectives were considered and how they were balanced between growth and adjustment objectives.

12. The MONA database includes information on what were the specific objectives of each IMF-supported program based on program documents presented to the IMF Executive Board. The MONA classifies program objectives into 17 categories in total, and IMF-supported programs typically involve multiple objectives. In this paper, for analytical purposes, program objectives over 17 categories are grouped into three broad categories of Adjustment, Growth, and Vulnerability Management (Vulnerability in short). Program objectives within each broad category are further classified into three areas of Fiscal, Financial and Other (Table 2). In the discussion below, program objectives in the Adjustment and Vulnerability categories are considered to reflect primarily sustainability considerations, while those in the Growth category represent growth considerations.

[3] Appendix II illustrates a simple analytical framework that explains the determination of adjustment and growth in program design and has guided the empirical analysis in this report.

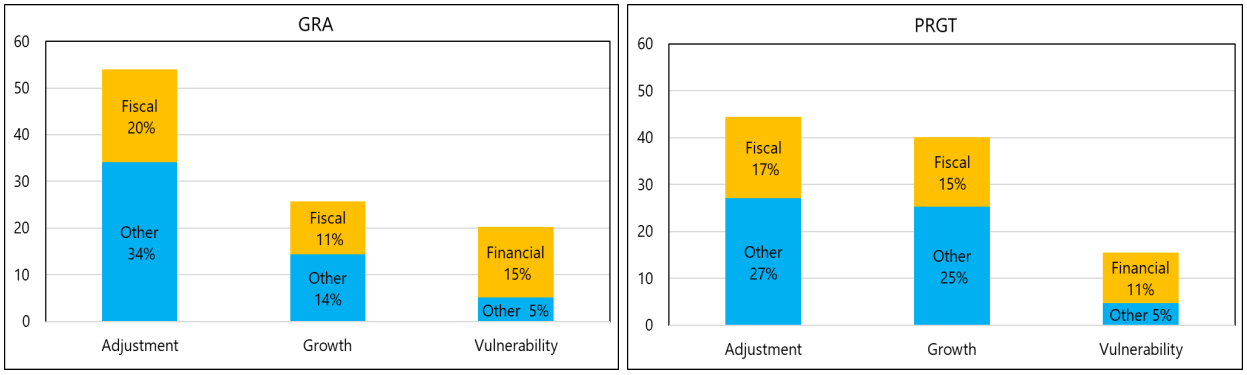
Table 2. Classification of Program Objectives

	Adjustment	Growth	Vulnerability	Total
Fiscal	Fiscal Revenue Public Expenditure	Social Enterprise Governance		5
Financial			Central Bank Financial Sector	2
Other	Macroeconomic External Stability Monetary Inflation Exchange Rate	Economic Growth Poverty Reduction Trade Pro-Growth	Others	10
Total	7	7	3	17

Sources: MONA database; IEO staff calculations.

13. Figure 2 shows the composition of program objectives of GRA and PRGT programs based on the classification in Table 2. The share of growth objectives is significantly higher in PRGT programs (40 percent) than in GRA programs (25 percent), while the opposite is the case for the shares of adjustment and vulnerability objectives. This notable difference in the share of growth objectives is broadly consistent with the notion that adjustment features more prominently than growth promotion in GRA programs while adjustment and growth objectives are more balanced in PRGT programs.

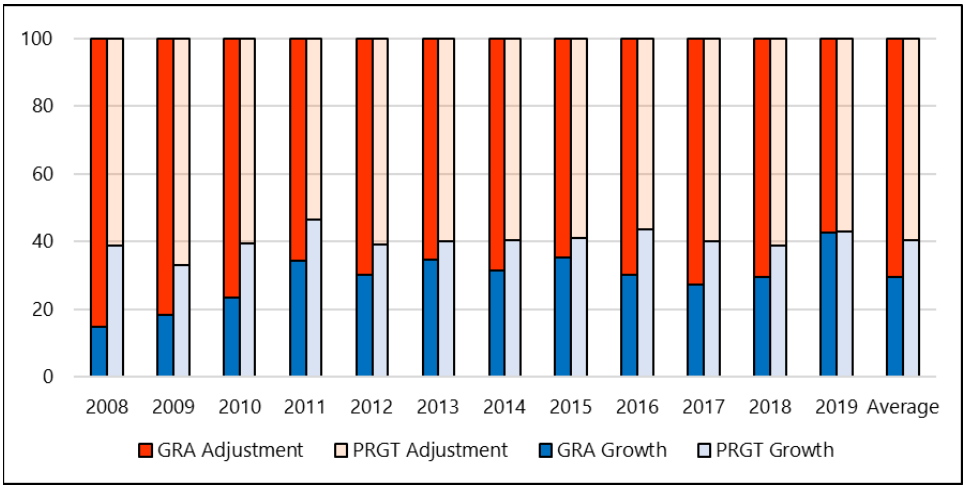
Figure 2. Composition of Program Objectives
(In percent)



Sources: MONA database; IEO staff calculations.

14. Figure 3 shows how the composition of program objectives has changed during the period of 2008–19, where the adjustment category is now broadly defined to include both adjustment and vulnerability objectives. It is notable that in GRA programs the share of growth objectives increased rather steeply to about 35 percent in programs approved in 2011, up from 24 percent in 2010, and has since remained broadly at the increased level, reflecting in part the IMF's increased attention to growth. For PRGT programs, the share of growth objectives has been stable over the sample period.

Figure 3. Share of Growth and Adjustment Objectives: 2008–19
(In percent)

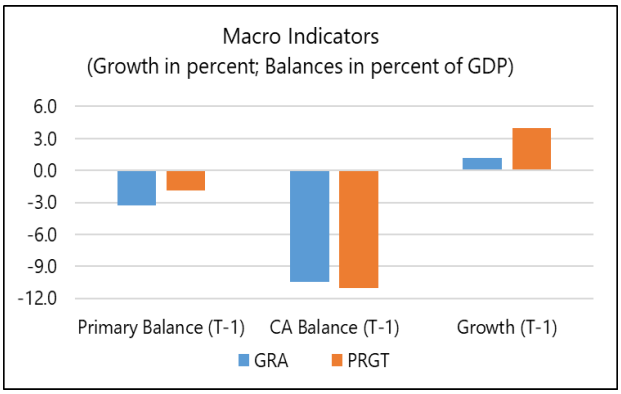


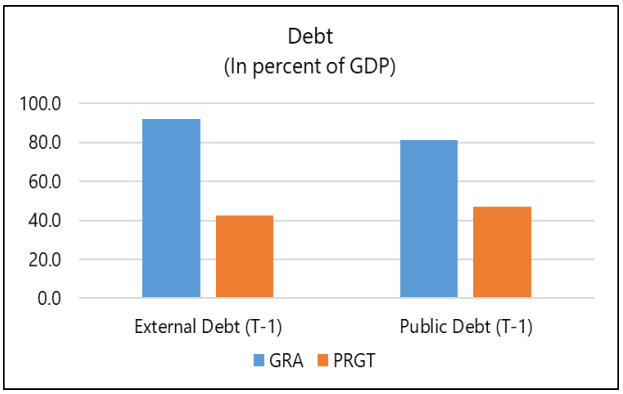
Sources: MONA database; IEO staff calculations.
Note: Adjustment category is defined broadly to include vulnerability-related objectives. For each year, the left (right) bar is for GRA (PRGT) programs approved in that year.

Initial Conditions

15. Initial conditions appear in general to have been better for PRGT programs than GRA programs in terms of both flow and stock indicators. PRGT programs had lower primary deficits, higher growth in the pre-program year, and much lower external and public debt ratios (Figure 4).[4] The initial current account deficit averaged slightly higher in PRGT than in GRA cases.

Figure 4. Initial Conditions





Sources: WEO database; IEO staff calculations.

[4] A more consistent comparison of initial debt ratios between GRA and PRGT programs would require consideration on country-specific debt carrying capacity, which is beyond the scope of this paper.

Time Pattern of Growth and Adjustment

16. In GRA programs, growth outcomes exhibit a U-shaped trajectory with the trough in the first year of the program (T) followed by a rapid recovery in growth in the next year and a more modest recovery in the subsequent periods (Figure 5a). Notable is the wide range of growth outcomes during the first year of the program, as indicated by the quartile range in shade. Growth outcomes consistently underperform growth projections (at program approval), indicating optimism bias embedded in initial program projections. Within GRA programs, notable differences are found between crisis and other programs (which include some programs in response to home-grown BOP crises) (Figure 5b). The U-shaped pattern in growth trajectories is far more pronounced in crisis programs than in other programs. The negative interquartile range in year T in Figure 5a is accounted for almost entirely by crisis programs.[5] Other GRA programs exhibit a steady recovery in growth in both projections and outcomes while, in crisis programs, an inverted U-shaped pattern is more visible in both growth outcomes and projections during and after the program. Growth outcomes of other programs show relatively small cross-country variation as indicated by the relatively narrow interquartile range.

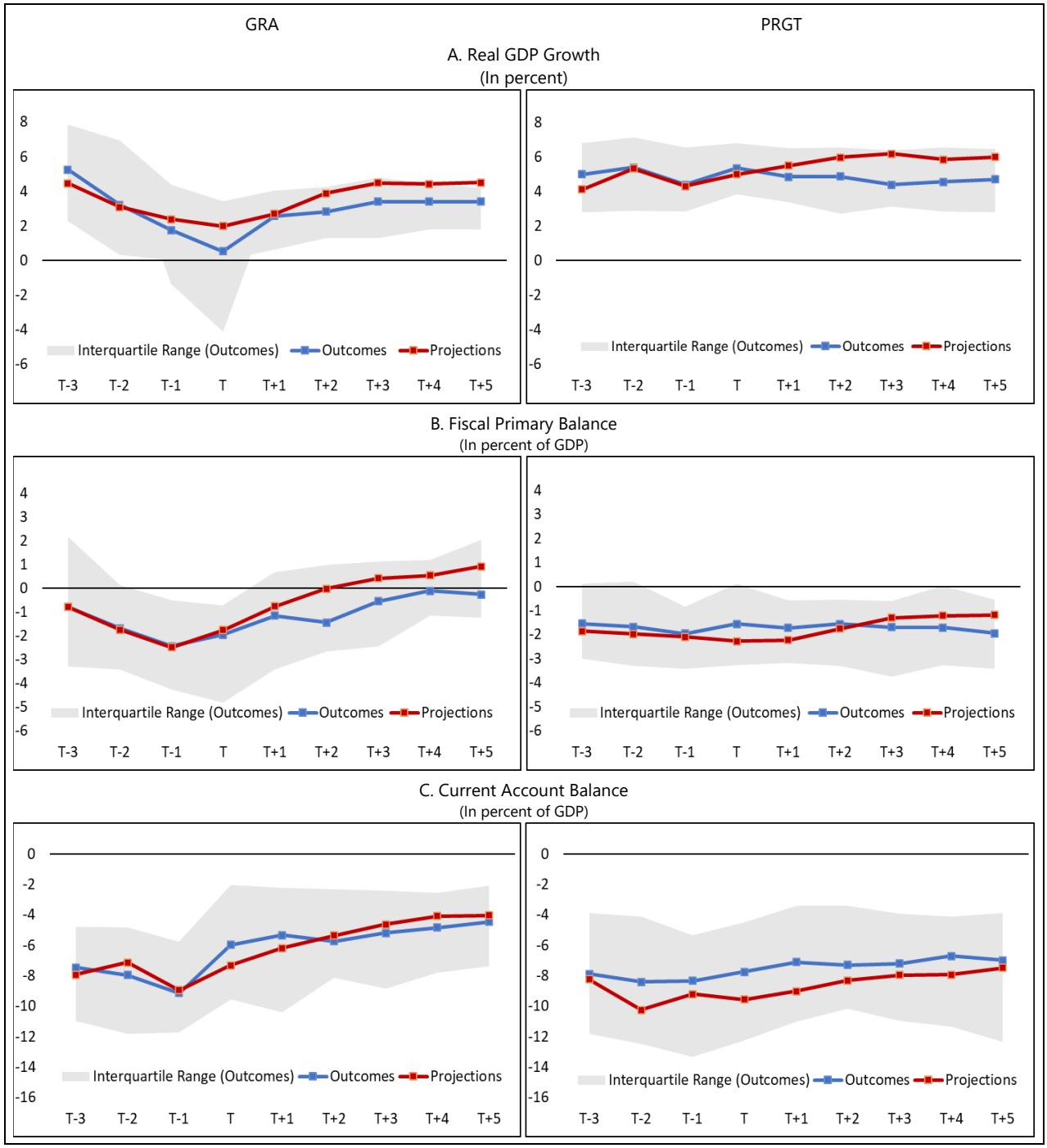
17. In PRGT programs, there is less pattern in the trajectory of growth outcomes. It is notable that an initial modest recovery in growth at T is followed by a steady decline in growth until T+3 before leveling off (Figure 5a). In contrast, growth projections show the opposite time pattern with a steady recovery in growth until T+3 before being reversed. As in GRA programs, growth outcomes generally fall short of projections (except for year T), suggesting optimism bias in initial growth projections.

18. Like growth trajectories, both fiscal and current account (CA) balances exhibit a U-shaped pattern in GRA programs as in the case of growth trajectories, but the trough is in year T-1 in case of CA balance (Figure 5a, Panels B and C). Such a U-shaped pattern is far less visible in PRGT programs where the trajectories of fiscal and CA outcomes are quite stable over time with little improvements. GRA programs show on average smaller fiscal and CA deficits in outcomes and projections than PRGT programs. In GRA programs, fiscal outcomes underperform projections by significant margin while CA outcomes are broadly consistent with projections. In PRGT programs, fiscal outcomes fall short of projections in the later years of the program or after the program while CA outcomes consistently outperform projections. It is notable, however, that CA outcomes vary widely across PRGT programs as indicated by the quite large interquartile range.

19. The U-shape pattern observed for fiscal and CA balances in GRA programs is primarily driven by crisis programs where fiscal outcomes significantly underperform projections (Figure 5b). Fiscal outcomes improve rather steadily in both crisis and other GRA programs but underperform projections by larger margin in crisis programs. CA outcomes are less steady in both crisis and other programs with initial improvements being reversed later.

[5] About 41 percent of GRA programs in the sample experienced real GDP contraction (i.e., negative growth) in the first year of the program, 65 percent of which are accounted for by crisis programs.

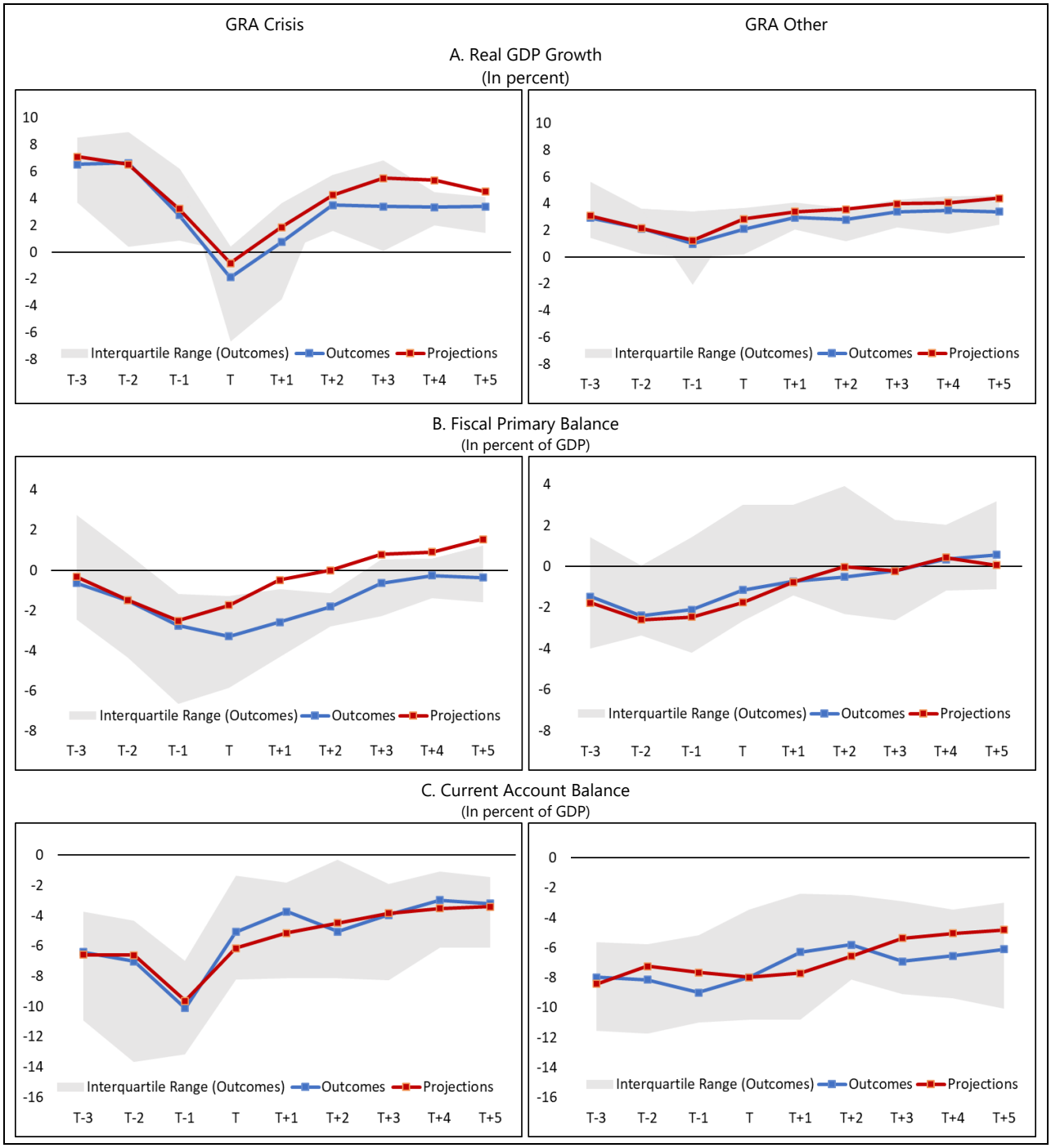
Figure 5a. Growth and Adjustment Trajectories: GRA and PRGT Programs (Cross-country medians)



Sources: WEO database; IEO staff calculations.

Note: All projections refer to initial program projections made at program approval. Outcomes and projections represent cross-country medians. Data availability is not uniform across periods mainly because post-program outcome data are not yet available for recently completed programs. Due to the presence of successor programs for the same countries in the sample, there is overlap in the data presented over the period and, therefore, the results are not always fully consistent with those based on program periods only.

Figure 5b. Growth and Adjustment Trajectories: Crisis and Other GRA Programs (Cross-country medians)



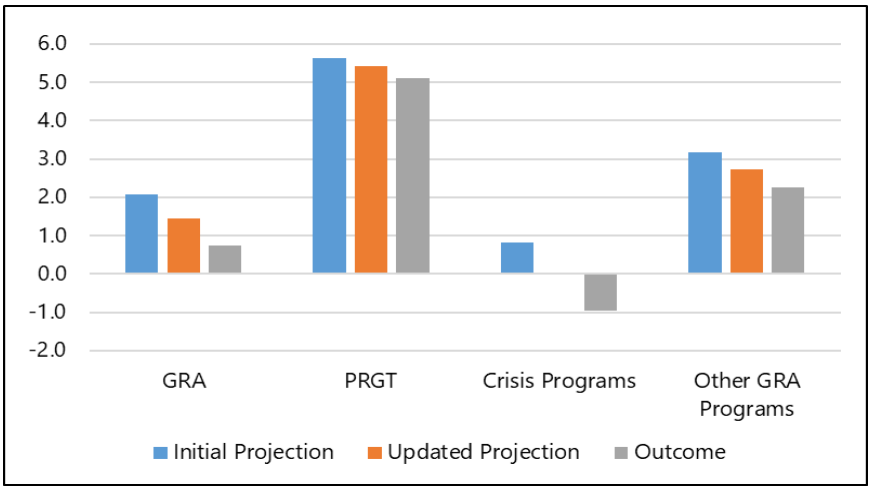
Sources: WEO database; IEO staff calculations.

Note: All projections refer to initial program projections made at program approval. Outcomes and projections represent cross-country medians. Data availability is not uniform across periods mainly because post-program outcome data are not yet available for recently completed programs. Due to the presence of successor programs for the same countries in the sample, there is overlap in the data presented over the period and, therefore, the results are not always fully consistent with those based on program periods only.

Growth Outcomes Relative to Projections

20. In both GRA and PRGT programs, growth projections were revised downwards over the program period, generally more so in GRA programs and particularly in crisis programs than in PRGT programs (Figure 6).[6] Optimism bias in initial program projections is on the order of 1.3 percentage points in GRA programs (1.8 percentage points and 0.9 percentage points in crisis and other programs, respectively) while it is relatively small at 0.5 percentage points in PRGT programs (Table AVI.1, Appendix VI). Updated (one-year-ahead) program projections were typically more realistic than initial projections in both GRA and PRGT programs.

Figure 6. Growth: Projections and Outcomes
(In percent; cross-country averages)

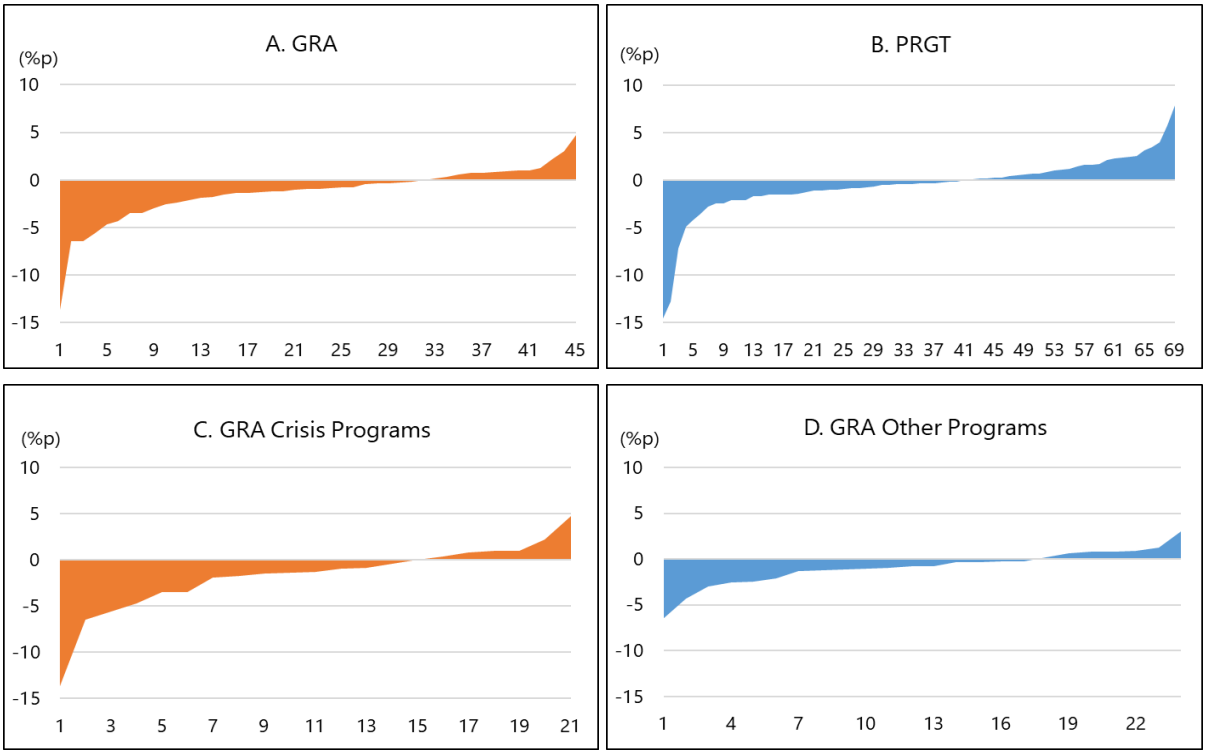


Sources: WEO database; IEO staff calculations.
Note: Cross-country averages are calculated based on annual averages of each program in the sample. Program periods used to calculate annual averages differs between initial projections and outcomes/updated projections.

21. Growth outcomes relative to initial program projections are widely dispersed in both GRA and PRGT programs (Figure 7). For GRA programs, growth fell short of projections by more than an annual average of ½ percentage points in 58 percent of cases; more than 1 percentage point in 47 percent of cases; and in 25 percent of cases the growth shortfall was greater than 2.2 percentage points (Table 3). Within GRA programs, growth shortfalls were on average larger in crisis programs than in other GRA programs. For PRGT programs, growth shortfalls were larger than ½ percentage points in 42 percent of cases and larger than 1.5 percentage points in 25 percent of cases.

[6] In the remainder of this chapter, the empirical analysis is based on the data for program periods only. Thus, the results could differ from those in Figures 5a and 5b (see the note to these figures). For consistent comparison between program outcomes and projections, the program sample is limited to 114 programs for which both projection (initial and updated) and outcome data are available for one year or longer.

Figure 7. Distribution of Growth Outcomes Relative to Initial Projections (In percentage points)



Sources: WEO database; IEO staff calculations.
Note: Data represent growth deviations (actual minus projection) in percentage points.

Table 3. Growth Outcomes Relative to Initial Projections by Program Type

Program	Min[1]	Max[1]	Share of Programs with Growth Shortfall > 0.5%p[2]	Share of Programs with Growth Shortfall > 1.0%p[2]	Bottom 25% Threshold[1]
ALL	-14.6	7.8	48.3	38.6	-1.6
GRA	-13.6	4.8	57.8	46.7	-2.2
Crisis programs	-13.6	4.8	61.9	52.4	-3.5
Other GRA programs	-6.5	3.0	54.2	41.7	-1.6
PRGT	-14.6	7.8	42.0	33.3	-1.5

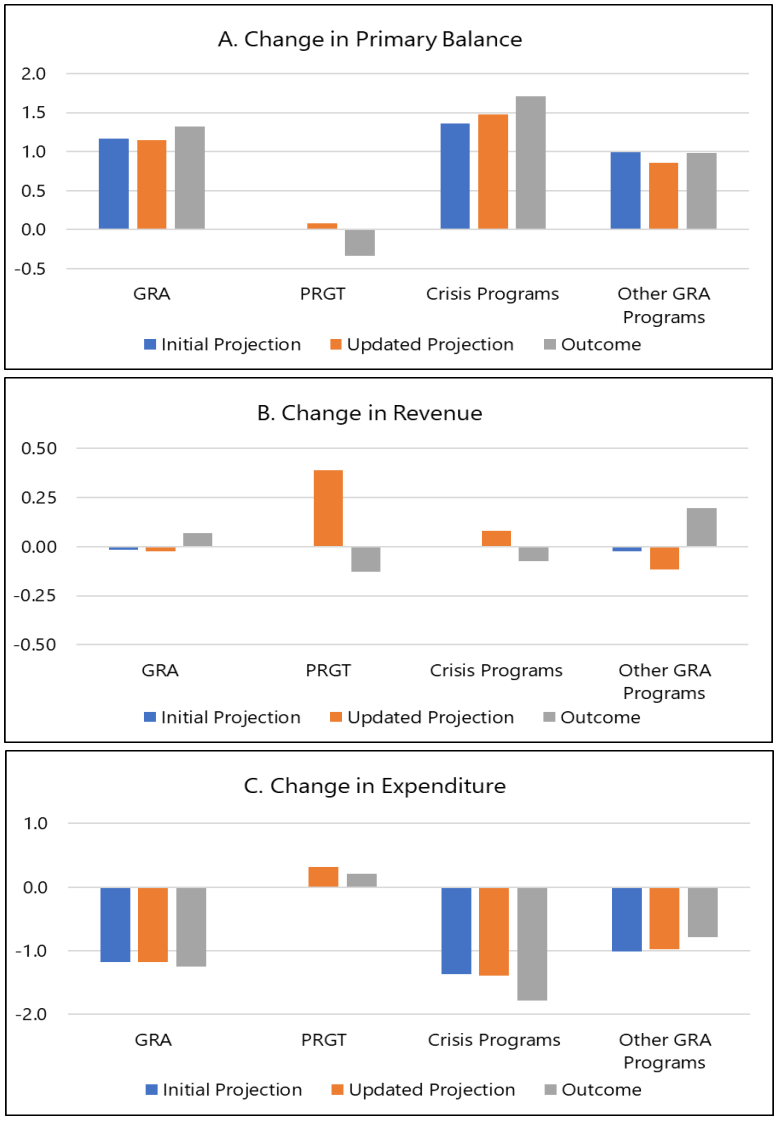
Source: WEO database; IEO staff calculations.
[1] The numbers refer to growth deviations (actual minus projection) in percentage points.
[2] In percent.

Adjustment Outcomes Relative to Projections

22. ***Fiscal adjustment***. In contrast to growth projections, fiscal projections were on average unbiased in GRA programs with only minor difference from fiscal outcomes while modestly optimistic in PRGT programs (Figure 8). Within GRA programs, differences between fiscal

outcomes and projections are on average larger in crisis programs where fiscal outturns were stronger than projected particularly in expenditure adjustment. Programmed fiscal adjustment was significant on the order of 1.2 percent of GDP in GRA programs on an annual average basis but tiny in PRGT programs (see Table AVI.1 in Appendix VI).[7] The composition of fiscal adjustment was dominated by expenditure adjustment in both projections and outcomes in GRA programs while more even in PRGT programs.

Figure 8. Fiscal Adjustment: Projections and Outcomes (In percent of GDP; cross-country average)

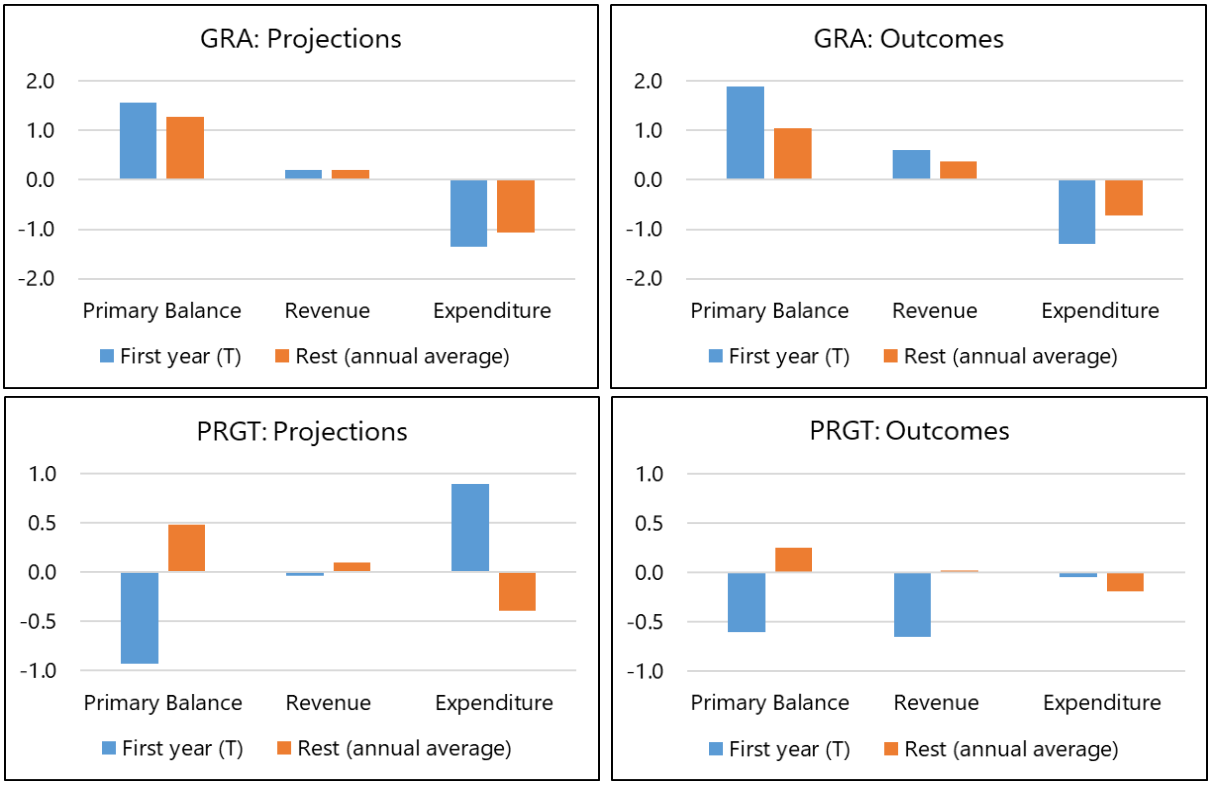


Sources: WEO database; IEO staff estimates.
Note: Cross-country averages are calculated based on annual averages of each program in the sample. Program periods used to calculate annual averages differ between initial projections and outcomes/updated projections.

[7] The sample standard deviation of programmed fiscal adjustment is also smaller for PRGT programs (1.6 percent of GDP) than in GRA programs (2.5 percent of GDP).

23. In terms of phasing, fiscal adjustment was front loaded in GRA programs (more so on the expenditure side and in program outcomes) while back loaded in PRGT programs with fiscal easing in the first year of the program followed by fiscal tightening in the rest of the program period (Figure 9).

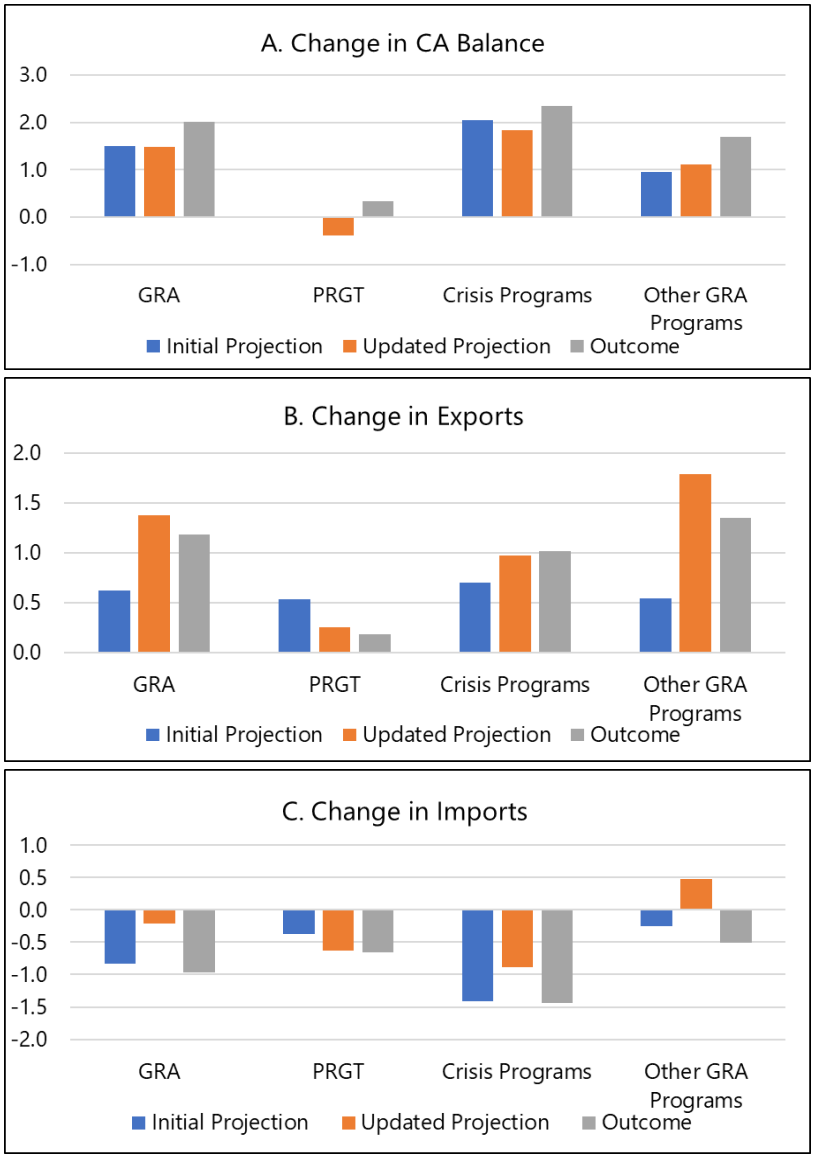
Figure 9. Phasing of Fiscal Adjustments: Projections and Outcomes
(In percent of GDP; cross-country average)



Sources: WEO database; IEO staff estimates.
Note: The sample includes programs with program duration of two years or longer only and is therefore smaller than the sample used for Figure 8.

24. ***External adjustment***. In GRA programs, programmed CA adjustment averaged at 1.5 percent of GDP while outturns were on average 2.0 percent of GDP, largely aided by stronger export performance than projected (Figure 10; see Table AVI.1 in Appendix VI). Within GRA programs, both programmed and actual CA adjustments were stronger and relied more on import compressions in crisis programs than in other programs where improved exports played a greater role than import compression. Programmed and actual CA adjustments were both far smaller in PRGT programs relative to GRA programs, but subject to large cross-program variation ranging from -10.7 percent to 6.3 percent of GDP in projection and -15.7 percent to 9.7 percent of GDP in outturns (see Table AVI.1 in Appendix VI).

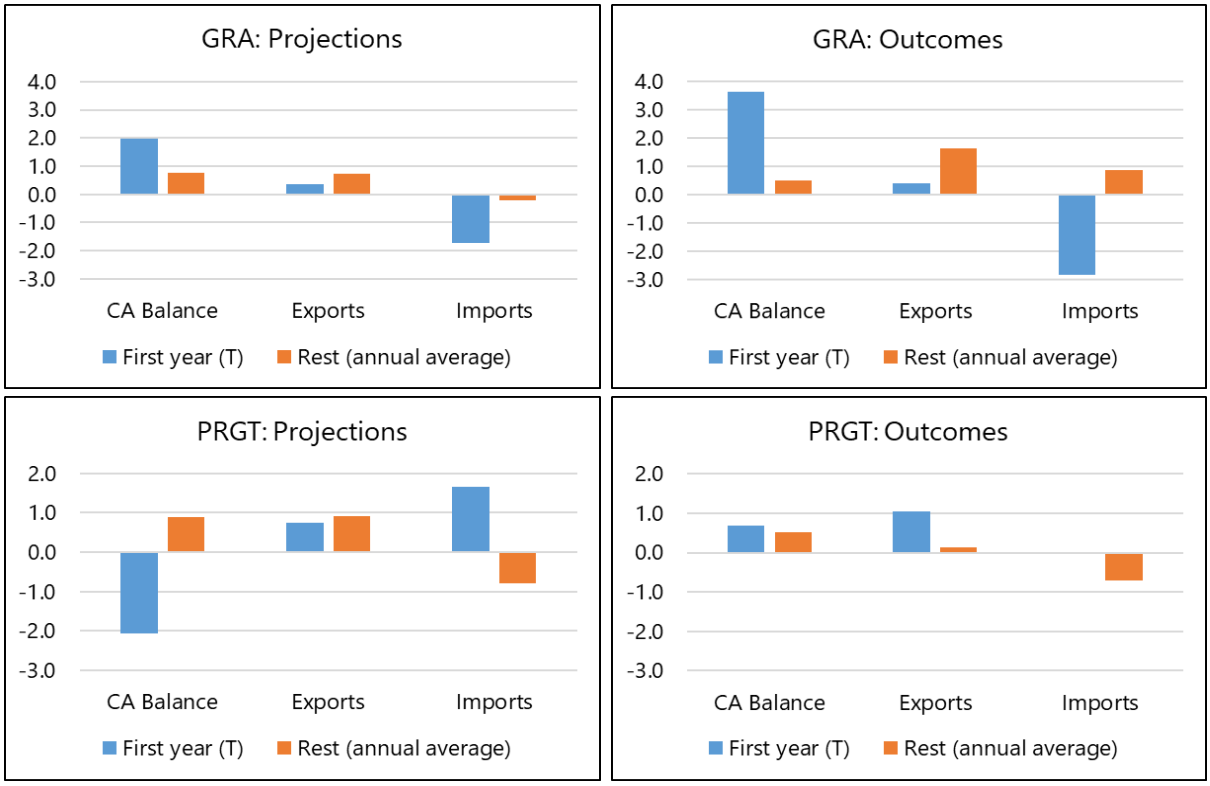
Figure 10. External Adjustment Projections and Outcomes
(In percent of GDP; cross-country average)



Sources: WEO database; IEO staff estimates.
Notes: Cross-country averages are calculated based on annual averages of each program in the sample. Program periods used to calculate annual averages differ between initial projections and outcomes/updated projections.

25. In terms of phasing, programmed CA adjustment was front loaded in GRA programs but back loaded in PRGT programs, which is broadly in line with the phasing pattern for fiscal adjustment (Figure 11). In case of GRA programs, front loading is even more pronounced in program outcomes largely driven by import compression. In sharp contrast to program projection, actual CA adjustment in PRGT programs was evenly phased largely because projected increase in imports in the early phase of the program did not materialize.

Figure 11. Phasing of External Adjustments
(In percent of GDP)



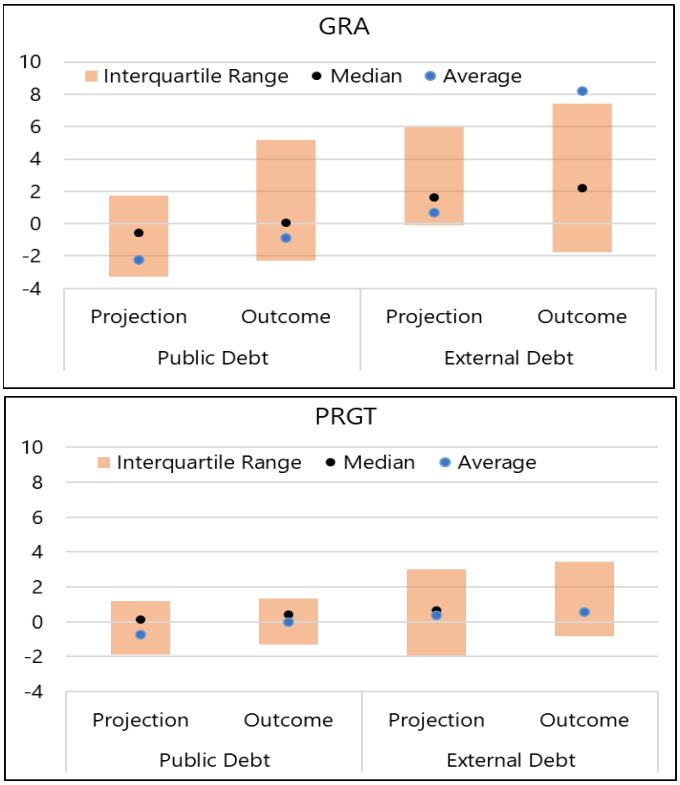
Sources: WEO database; IEO staff estimates.
Note: The sample includes programs with program duration of two years or longer only.

26. ***Change in debt***. Both GRA and PRGT programs targeted debt reduction on the order of 0.7–2.2 percent of GDP for public debt and 0.0–0.9 percent of GDP for external debt on an annual average basis (see Tables AVI.1 in Appendix VI). But both public and external debt have on average increased, rather than decreased, in both GRA and PRGT programs. Moreover, there is large cross-country variation in debt projections and outcomes especially in GRA programs as indicated by large interquartile ranges (Figure 12).

Decomposition of BOP Need

27. Drawing on the 2018 ROC, the counterfactual BOP need, which is defined specifically as the BOP need assuming continuation in the pre-program balance of payments, is estimated for 40 select programs (22 GRA and 18 PRGT) arranged for 17 countries covered in the case studies prepared for the evaluation. The estimation and decomposition of the BOP need is undertaken following the same methodology used by the 2018 ROC. Appendix III explains the methodology in greater detail and discusses some additional results on the BOP need decomposition.

Figure 12. Distribution of Change in Debt
(In percent of GDP; annual average)



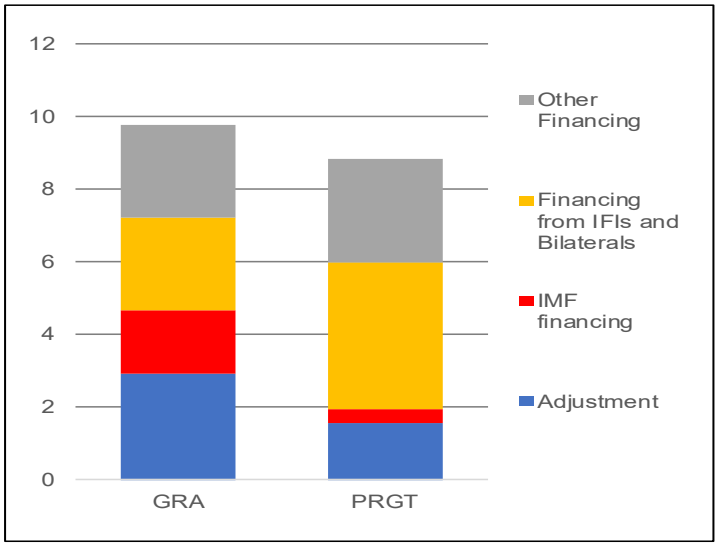
Source: IEO staff calculations.

28. The decomposition results are summarized in Figure 13. The annualized BOP need is on average larger in GRA programs (9.8 percent of GDP) than in PRGT programs (8.8 percent of GDP), and so are programmed adjustment and Fund financing. The overall results in Figure 13 are consistent with the findings of the 2018 ROC—namely, (i) both CA adjustment (relative to counterfactual CA deficits) and Fund financing are significantly larger for GRA programs than PRGT programs, (ii) Fund financing accounts for a much smaller portion in PRGT programs, (iii) program design envisages substantially stronger contributions from other official creditors in PRGT programs than in GRA programs, and (iv) both GRA and PRGT programs are expected to catalyze increased financing from other sources.

29. The bivariate relationship between CA adjustment (relative to counterfactual CA deficits) and program growth projection is negative and statistically significant, suggesting a short-run tradeoff between CA adjustment and growth envisaged in initial program design (Figure 14). This result is interesting for two reasons. First, programmed CA adjustment could be associated either positively or negatively with growth in the short run depending on the balance in adjustment between exports and imports. Second, the methodology used to calculate CA adjustment relative

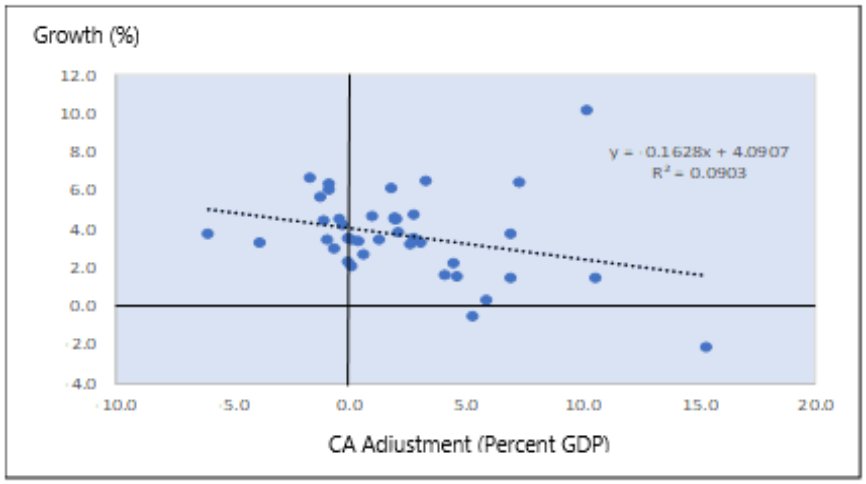
to counterfactual CA deficits tends to generate positive rather than negative association, if any, between CA adjustment and growth projection.[8]

Figure 13. BOP Need Decomposition: Select Programs (In percent of GDP; annual average)



Source: IEO staff estimates.

Figure 14. Programmed CA Adjustment and Growth Projection (Select programs; program-period average)



Source: IEO staff estimates.

[8] Counterfactual CA deficits (CADs) were constructed by applying the CAD/GDP ratio at T-1 to projected GDP over the program period (see Appendix III). As a result, higher growth projections over the program period would ceteris paribus increase counterfactual CADs and hence CA adjustment in the BOP decomposition, suggesting a positive relationship, if any, between CA adjustment and growth projection.

IV. Sustainability and Growth Considerations in Initial Program Design

30. This section discusses the results of cross-country analysis about how sustainability and growth considerations were incorporated in initial program design, particularly in the design of fiscal policy. In the analysis, sustainability and growth consideration are assessed through the lens of programmed fiscal policy because fiscal adjustment is at the center of macroeconomic stabilization in the program context and relates to policy instruments under control of country authorities. The cross-country analysis in this section is complementary to qualitative assessment of the attention paid to growth in program design discussed in select case studies prepared for the evaluation.

31. The approach used for assessment is guided by the analytical framework developed by Bohn (1998, 2008) and used in subsequent research on debt sustainability and fiscal space (Mendoza and Ostry, 2008). Specifically, the analytical framework suggests that a sufficient condition for the government to satisfy its intertemporal budget constraint is that the primary balance reacts positively to lagged debt. This criterion could be considered as a weak condition for debt sustainability. Using the framework, Ostry and others (2010) further developed a stronger sustainability criterion based on a nonlinear fiscal reaction function which incorporates the notion of fiscal fatigue and applied the criterion to determine country-specific debt limits and fiscal space of advanced economies.

32. The same analytical framework is used for the cross-country analysis in this section. To be specific, a fiscal reaction function, which relates programmed fiscal policy to lagged debt ratio and the output gap, is estimated and the estimated reaction coefficients are used as a basis for assessing how sustainability and growth considerations were reflected in the design of fiscal policy.

Fiscal Reaction Function

33. According to the debt sustainability criterion developed in the literature, sustainability considerations would call for a positive response of the fiscal primary balance to the lagged public debt ratio—at least over a certain range of debt ratios if not for the full range. Growth considerations would suggest on average a positive response of the primary balance to the output gap (defined as actual GDP minus trend GDP) so that fiscal policy tends to be counter-cyclical in nature with tendency for fiscal easing (tightening) at times of weak (strong) economic activities.

34. To model such fiscal policy response, the analysis considers the following baseline specification for the fiscal reaction function which is given by

$$PBY_t = \beta_0 + \beta_1 \cdot PDY_{t-1} + \beta_2 \cdot YGAP_t \quad (1)$$

where PBY, PDY and YGAP refer to the fiscal primary balance as a share of GDP, the public debt-to-GDP ratio and the output gap (defined as actual GDP minus trend GDP and measured in percent of trend GDP), respectively. As discussed above, sustainability considerations would

imply $\beta_1 > 0$, and higher values of β_1 would generally indicate stronger sustainability considerations. Growth considerations would be captured by $\beta_2 > 0$ with larger positive values of β_2 reflecting stronger growth considerations.

Data and Estimation Results

35. The fiscal reaction function specified above is estimated by using a panel sample of initial projection data taken from the first WEO vintage published after program approval. The output gap (YGAP) is constructed by using a simple log-linear trend model for real GDP. Specifically, the log-linear trend is estimated for real GDP over the 10-year period prior to the program and then extrapolated into the program period. The output gap is constructed as the difference between projected real GDP and trend real GDP and measured in percent of trend real GDP.

36. Table 4 presents the estimation results. The results in columns (1) and (3) based on the linear specification suggest that both sustainability and growth considerations were well reflected in programmed fiscal policy in GRA programs but less clearly so in PRGT programs. The reaction coefficients are of the expected sign and statistically significant in GRA programs while none is significant in PRGT programs. It is also notable that the simple linear specification explains about one-third of the sample variation in the GRA sample, while explaining very little in the PRGT sample.

Table 4. Fiscal Reaction: Initial Program Design

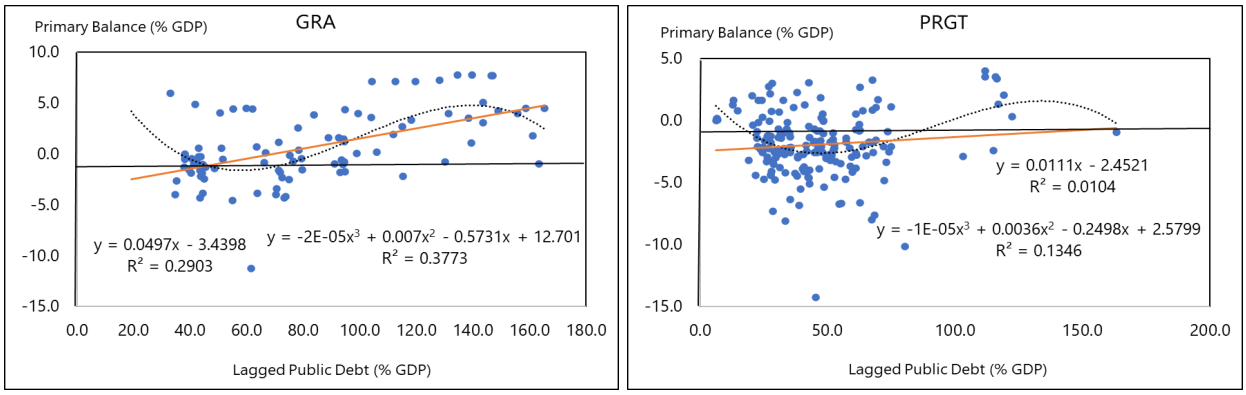
	GRA		PRGT	
	(1)	(2)	(3)	(4)
	PBY	PBY	PBY	PBY
YGAP	0.092**	0.108**	-0.002	0.019
	(0.043)	(0.042)	(0.014)	(0.015)
PDY_lagged	0.051***	-0.661***	0.013	-0.227***
	(0.008)	(0.193)	(0.011)	(0.079)
PDY_lagged_square		0.008***		0.003*
		(0.002)		(0.002)
PDY_lagged_cubic		-2.6e-05***		-9.5e-06
		(7.7e-06)		(8.6e-06)
Constant	-2.859***	15.959***	-2.51***	2.359**
	(0.777)	(4.875)	(0.486)	(1.094)
No observation	90	90	182	182
R^2	0.337	0.45	0.011	0.15
SE of regression	2.908	2.679	2.624	2.447

Source: IEO staff estimates.
Note: Standard errors are in parenthesis. *p<0.1, **p<0.05, ***p<0.01.

37. Following Ghosh and others (2013), the analysis also considers a nonlinear fiscal reaction function. Figure 15 suggests that the bivariate relationship between the primary balance and lagged debt ratio could be highly nonlinear. The nonlinear trend lines in the figure suggest a positive response of the primary balance to lagged debt ratio over the interval of debt ratio

between 50 percent and 140 percent of GDP in both types of programs. An important implication of nonlinear fiscal reactions is that the positive fiscal reaction is reduced when debt ratio exceeds a certain threshold. This feature in fiscal outcomes is attributed to fiscal fatigue in Ghosh and others (2013). In the context of program design, it could reflect some feasibility constraints in fiscal adjustment or growth considerations beyond what is captured by the reaction to the output gap.

Figure 15. Primary Balance and Lagged Public Debt: Initial Program Design



Source: IEO staff estimates.

38. The results of additional regressions in columns (2) and (4) of Table 4 provide good support for the nonlinearity found in the data, particularly in the GRA sample. Notably, the introduction of nonlinearity improves the goodness of fit significantly in both GRA and PRGT samples and finds a significant relationship between the fiscal primary balance and the (lagged) public debt ratio in GRA as well as PRGT programs, albeit less strong in the latter. While the nonlinear fiscal reaction may capture some growth considerations, the direct response to the output gap continues to be small and not statistically significant in PRGT programs, suggesting that initial program design may have assumed automatic stabilizers to be weak at best in low-income countries (LICs), consistent with the existing evidence.

39. Taken together, these findings suggest that in GRA programs, sustainability and growth considerations were both at play in fiscal policy envisaged in initial program design. For PRGT programs, however, the evidence is weaker for sustainability and growth considerations in initial fiscal program design.

V. Sustainability and Growth Considerations in Program Adaptation

40. Program projections in initial program design provide the most comprehensive snapshot of the macroeconomic framework calibrated to achieve targeted adjustment and growth given available financing. However, focusing only on initial program design would miss a crucial aspect of program design: the flexible adaptation of programs in response to interim macroeconomic outcomes in the context of periodic reviews of program implementation (Mussa and Savastano, 1999).

41. Focusing on this latter aspect, this section assesses sustainability and growth considerations in program adaptation by examining how programmed fiscal adjustments were modified over the program period in response to interim growth and adjustment outcomes relative to projections.[9] Although program reviews are typically undertaken on a semiannual frequency, an empirical analysis based on annual data still provides useful insights into how programs have been adapted to reflect incoming information on macroeconomic outcomes.

Fiscal Reaction Function

42. As in the previous section, assessment of sustainability and growth consideration in the context of program adaptation is based on a fiscal reaction function which models the programmed fiscal response to interim adjustment and growth outcomes relative to projections (i.e., adjustment and growth forecast errors). To this end, the analysis considers the following fiscal reaction function for program adaptation:

$$\Delta E(FA)_{t+1}(= E_t FA_{t+1} - E_{t-1} FA_{t+1}) = \gamma_0 + \gamma_1 FE(FA)_t + \gamma_2 FE(G)_t \quad (2)$$

where $E_s FA_z$ stands for program projections formulated at time *s* for fiscal adjustment (FA) at time z, and *G* is the real GDP growth rate. FE(FA) and FE(G) stand for one-period-ahead forecast errors of fiscal adjustment and growth, respectively. The dependent variable, $\Delta E(FA)_{t+1}$, represents the revision made in year t to the programmed fiscal adjustment in year t+1 (i.e., the difference between two-period-ahead and one-period-ahead projections for fiscal adjustment in year t+1).[10] This fiscal reaction function models how programmed fiscal adjustments were revised in each program year (after the first year of the program) in response to lagged forecast errors of fiscal adjustment and growth.

43. The reaction coefficients, γ_1 and γ_2, provide the basis of our assessment of sustainability and growth considerations in fiscal policy in the context of program adaptation. As discussed below, it should be noted that each reaction coefficient captures both sustainability and growth considerations. Sustainability considerations would require $\gamma_1 < 0$. In this case, adjustment shortfalls (i.e., $FE(FA) < 0$) in the current year lead to stronger fiscal adjustment than otherwise in the next year. At the same time, positive adjustment surprises (i.e., $FE(FA) > 0$) lead to weaker fiscal adjustment and *ceteris paribus* higher growth than otherwise, providing some support for

[9] In a similar context, IEO (2003) analyzed how revisions in fiscal targets during the program are linked to contemporaneous changes in growth projections.

[10] To be specific, consider for example a three-year program over the period from T to T+2. For fiscal adjustment in year T+2, there are two program projections: (i) one-year-ahead projection formulated in year T+1 and (ii) two-year-ahead projection formulated in year T. The dependent variable for year T+1 is the difference between these two projections for fiscal adjustment in T+2. The corresponding independent variables are one-year-ahead forecast errors of adjustment and growth observed in year T+1, which are measured by the difference between actual values in T+1 and projections made in year T.

growth considerations.[11] Growth considerations would require $\gamma_2 > 0$, in which case growth shortfalls (i.e., $FE(G) < 0$) in the current year result in a reduction in planned fiscal adjustment and thus higher growth than otherwise in the next year. On the other hand, positive growth surprises (i.e., $FE(G) > 0$) imply stronger fiscal adjustment than otherwise in the next year, easing sustainability concerns if any.

44. For given reaction coefficients and forecast errors, the correlation between adjustment and growth forecast errors would matter for the average magnitude of the revisions in programmed fiscal adjustments. Suppose, for example, that $\gamma_1 < 0$ and $\gamma_2 > 0$. If adjustment and growth forecast errors are positively correlated, fiscal reactions to adjustment and growth forecast errors are likely to offset each other to some extent given the opposite sign of the reaction coefficients, resulting in more muted fiscal responses to interim forecast errors. Conversely, fiscal reactions to forecast errors would likely reinforce each other if forecast errors are negatively correlated, resulting in larger revisions in programmed fiscal adjustment than otherwise.

Data and Estimation Results

45. In estimating the fiscal reaction function, fiscal adjustment (FA) is measured by the change in the fiscal primary balance/GDP (ΔPBY). To approximate the information available to country authorities and Fund staff at the time of projections, both actual and projection data are taken from various WEO vintages corresponding to program years. For instance, forecast errors are measured by using real time data as recorded in the WEO vintages matched with program years, rather than the latest actual data from the most recent WEO vintage.[12]

46. Table 5 presents sample statistics for growth and fiscal adjustment forecast errors in GRA and PRGT programs estimated for our data set. Program growth projections show a minor optimism bias in both GRA and PRGT programs—the sample mean of growth forecast errors is relatively small at -0.2 percentage points and -0.3 percentage points in GRA and PRGT programs, respectively. Growth shortfalls (negative growth forecast errors) have been relatively more frequent in PRGT programs (56 percent for GRA vs. 60 percent for PRGT). Fiscal adjustment has on average fallen somewhat short of programmed in GRA programs (by 0.2 percentage points) but has in fact slightly exceeded programmed amounts in PRGT programs as fiscal adjustment shortfalls have been relatively more frequent in GRA programs (54 percent for GRA and 46 percent for PRGT).

[11] Weaker fiscal adjustment in response to adjustment surprises is less likely to raise sustainability concerns unless γ_1 < -1 in which case adjustment surprises in the current year is more than fully offset in the next year and, therefore, cumulative fiscal adjustment over current and next years would on average be less than initially projected. Similarly, stronger fiscal adjustment in response to growth surprises is less likely to raise growth concerns unless $\gamma_2 > 1/m$ where m stands for fiscal multiplier, in which case cumulative growth would on average fall short of initial projection.

[12] The data for $E_t FA_{t+1}$ and $E_{t-1} FA_{t+1}$ are taken from the October WEO vintages of year t and year t-1, respectively. Forecast errors for year t are constructed based on *real time* data taken from the October WEO vintage of year t and projection data taken from the October WEO vintage of year t-1. The estimation sample is smaller than that used for the analysis of initial program design in Section III.A, largely because many programs went off track and the data used in regressions start from the second year of the program.

Table 5. Fiscal Adjustment and Growth Forecast Errors

Program Type	Forecast Error	N	Min	Median	Max	Mean	St dev
GRA	Growth[1]	79	-8.0	-0.2	5.5	-0.2	1.6
	Negative	44	-8.0	-0.9	0.0	-1.1	1.3
	Positive	35	0.0	0.6	5.5	1.0	1.2
	Fiscal Adjustment[2]	74	-12.9	-0.2	5.9	-0.3	2.4
	Negative	40	-12.9	-1.0	0.0	-1.6	2.3
	Positive	34	0.0	0.8	5.9	1.3	1.4
PRGT	Growth[1]	143	-5.3	-0.2	5.0	-0.3	1.3
	Negative	86	-5.3	-0.7	0.0	-1.0	1.1
	Positive	57	0.0	0.3	5.0	0.7	1.0
	Fiscal Adjustment[2]	142	-4.4	0.1	7.4	0.2	1.5
	Negative	65	-4.4	-0.5	0.0	-1.0	1.0
	Positive	77	0.0	0.8	7.4	1.1	1.2

Source: IEO staff estimates.
[1] In percentage points.
[2] In percentage points of GDP.

47. Sample variation (in terms of the standard deviation) in fiscal adjustment forecast errors is larger for GRA programs than for PRGT programs, suggesting that (one-year-ahead) fiscal projections were less accurate in GRA programs than in PRGT programs. Although not reported in Table 5, the sample correlation between growth and fiscal adjustment forecast errors is positive—implying that more adjustment has been associated with higher growth presumably reflecting the cyclical impact of growth on the fiscal accounts—but weak and statistically insignificant in both GRA and PRGT programs (0.24 for GRA and 0.1 for PRGT). Finally, Figure AVI.1 in Appendix VI shows scatter plots of the revisions in programmed fiscal adjustment (i.e., $E_t FA_{t+1} - E_{t-1} FA_{t+1}$) against (lagged) growth and adjustment forecast errors.

48. Table 6 presents the estimated fiscal reaction functions for GRA and PRGT programs. In addition to the baseline specification discussed above, shown in columns (1) and (3), an additional specification is considered in which forecast errors are disaggregated between positive and negative errors as shown in columns (2) and (4). Key findings from the regression results can be summarized as follows:

- First, the estimated reaction coefficients are of the expected sign and statistically significant in most cases.
- Second, sustainability considerations seem to have been stronger in GRA programs than in PRGT programs as indicated by the larger negative reaction coefficient γ_1 in the former (columns (1) and (3)), although the difference between GRA and PRGT programs is not statistically significant. The results in columns (2) and (4) reinforce this finding, and here the reaction coefficients of fiscal adjustment forecast errors are significantly different between GRA and PRGT programs. About 90 percent of adjustment shortfalls ($FE(FA) < 0$) were

programmed to be recovered in the next period in GRA programs while less than half of adjustment shortfalls were to be recovered in PRGT programs.

- Third, growth considerations seem to have on average played a stronger role in PRGT programs than in GRA programs as indicated by the larger positive reaction coefficient on $FE(G)$ in the former (columns (1) and (3)). The disaggregated results in columns (2) and (4) suggest both GRA and PRGT programs have paid attention to growth. Fiscal reaction to growth shortfalls ($FE(G) < 0$) in GRA programs is statistically significant and larger in magnitude than in PRGT programs. On the other hand, 87 percent of positive adjustment surprises ($FE(FA) > 0$) were to be reversed in the next period in PRGT programs which could indicate a greater attention to growth concerns, while only about 40 percent of positive adjustment surprises tended to be reversed in GRA programs.

Table 6. Fiscal Reaction: Program Adaptation

	GRA		PRGT	
	(1) $\Delta E(FA)_{t+1}$	(2) $\Delta E(FA)_{t+1}$	(3) $\Delta E(FA)_{t+1}$	(4) $\Delta E(FA)_{t+1}$
FE $(FA)_t$	-0.742***		-0.673***	
	(0.118)		(0.145)	
FE $(G)_t$	0.027		0.292***	
	(0.141)		(0.105)	
FE $(FA)_t < 0$		-0.905***		-0.424***
		(0.1)		(0.123)
FE $(FA)_t > 0$		-0.405**		-0.865***
		(0.195)		(0.246)
FE $(G)_t < 0$		0.264*		0.163
		(0.14)		(0.118)
FE $(G)_t > 0$		-0.293		0.480**
		(0.255)		(0.2)
Constant	0.054	-0.003	0.013	0.11
	(0.165)	(0.206)	(0.126)	(0.144)
N	74	74	142	142
R^2	0.619	0.663	0.321	0.341
SE (error term)	1.409	1.346	1.567	1.554

Source: IEO staff estimates.
Note: Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

49. Taken together, these findings suggest in broad terms that both sustainability and growth considerations were at play with respect to program adaptation of fiscal adjustment objectives. The evidence suggests that sustainability concerns were somewhat stronger in adapting GRA programs, while growth considerations may have played a stronger role in PRGT programs. The results for PRGT programs suggest that while attention to sustainability and growth was relatively limited in initial program design as discussed in Section IV, it strengthened over the program period via the process of program adaptation.

VI. Modeling Growth in Program Design and Adaptation

50. This section aims to assess how growth was modeled within the macroeconomic framework used in program design, both initially and when programs were adapted. To this end, a formal regression analysis is conducted to estimate the short-run relationship between adjustment and growth embodied in program projections by using initial and updated program projection data. It should be noted at the outset that the purpose of the regression analysis is not to establish a causal relationship between adjustment and growth, but to assess the underlying assumptions—particularly program assumptions on fiscal multipliers—used to formulate program projections. Section VII presents a similar analysis for the relationship between adjustment and growth outcomes and discusses the results in comparison with the findings in this section to assess the realism of the macro framework in program design.

51. Panel growth regressions use the following three sets of explanatory variables:

- *Adjustment parameters and external conditions* (program projections). Fiscal adjustment (ΔPBY), revenue adjustment (ΔGREVY), expenditure adjustment (ΔGEXPY), CA adjustment (ΔCABY), change in the public investment/GDP ratio (ΔPUBINVY) and change in the social spending/GDP ratio (ΔSOCIALY), as well as trading partners' growth (TPGR), and percentage change in the terms of trade (ΔTOT).[13]
- *Initial conditions* (real time data). Real GDP growth, the external debt/GDP ratio (EDY) and the public debt/GDP ratio (PDY) at T-1 (i.e., the year immediately prior to the program). The same initial conditions at T-1 were used in the regressions for both initial and updated program projections.
- *Country and program characteristics*. The country's institutional capacity at T-1 proxied by the government effectiveness index (GEI) published by the World Bank, and various dummies for program type and other country characteristics.

Estimation Results for Initial Program Projections

52. All projection data as well as *real time* data used in panel growth regressions are taken from the first WEO vintage published after program approval.

53. The regression results for GRA programs indicate a statistically significant short-run tradeoff between fiscal adjustment and growth embodied in initial program design (Table 7). The estimated coefficients of ΔPBY in columns (1)-(5) capture program assumptions on fiscal multiplier.[14] They are on the order of 0.37–0.51, which is broadly consistent with the broader

[13] PDY, ΔPUBINVY and ΔSOCIALY are included only individually for sensitivity check because their inclusion shrinks the sample substantially due to the limited data availability of these variables. See Tables 7 and 8 below.

[14] Strictly speaking, the estimated coefficient of ΔPBY would only be considered as a proxy for fiscal multiplier because ΔPBY represents fiscal adjustment measured by a change in the fiscal primary balance and not the change in the *structural* primary balance.

literature on fiscal multipliers (Gupta, 2020) and program assumptions discussed in several country case studies prepared for the evaluation. The estimated coefficients of ΔGREVY and ΔGEXPY in columns (6)-(10) suggest that GRA programs have on average assumed smaller fiscal multipliers for revenue-based adjustment than expenditure-based adjustment, which is again in line with existing evidence in the literature. In contrast to fiscal adjustment, no significant short-run tradeoff is detected between CA adjustment and growth projections in GRA programs, which is not surprising given the ambiguous theoretical relationship between them.

Table 7. Results of GRA Growth Regressions: Initial Program Projections

	GRA							
	(1) Growth	(2) Growth	(3) Growth	(4) Growth	(5) Growth	(6) Growth	(7) Growth	(8) Growth
ΔPBY	-0.372***	-0.471***	-0.513***	-0.041				
	(0.096)	(0.113)	(0.185)	(0.216)				
ΔGREVY					-0.244**	-0.393*	-0.348*	0.475
					(0.122)	(0.216)	(0.19)	(0.422)
ΔGEXPY					0.406***	0.335**	0.6***	0.432**
					(0.12)	(0.154)	(0.185)	(0.216)
ΔCABY	-0.156	-0.195	-0.132	-0.107	-0.155	-0.196	-0.238	-0.04
	(0.123)	(0.156)	(0.17)	(0.406)	(0.145)	(0.175)	(0.176)	(0.449)
TP_GR	0.972***	0.93***	1.144***	0.967***	1.036***	1.108***	1.206***	1.145***
	(0.162)	(0.238)	(0.17)	(0.188)	(0.142)	(0.307)	(0.158)	(0.199)
ΔTOT	0.04	0.086	0.071	0.14*	0.042	0.074	0.08	0.14**
	(0.042)	(0.06)	(0.049)	(0.072)	(0.048)	(0.062)	(0.056)	(0.065)
Growth (T-1)	-0.037	-0.006	-0.029	-0.055	-0.035	0.028	-0.024	-0.02
	(0.045)	(0.051)	(0.037)	(0.084)	(0.042)	(0.083)	(0.041)	(0.054)
EDY (T-1)	-0.003	-0.01	-0.001	-0.001	-0.001	0	0.001	0.001
	(0.005)	(0.008)	(0.005)	(0.008)	(0.005)	(0.01)	(0.005)	(0.007)
GEI (T-1)	-0.748*	-0.117	-1.025*	-0.702	-1.189***	-1.131	-1.269**	-0.209
	(0.396)	(0.575)	(0.561)	(0.99)	(0.447)	(0.821)	(0.618)	(0.844)
EA programs	-0.706*	-0.309	-0.532	0.888	-0.785**	-0.492	-0.641	1.159*
	(0.423)	(0.488)	(0.409)	(0.672)	(0.378)	(0.542)	(0.391)	(0.663)
Crisis programs	0.465	1.149	0.657	-1.116	0.74	0.522	0.672	-2.094**
	(0.604)	(1.112)	(0.615)	(1.23)	(0.556)	(1.41)	(0.586)	(1.05)
Small states	0.518	0.797	1.01	2.156*	1.564***	2.043	1.724***	2.709***
	(0.613)	(0.772)	(0.671)	(1.141)	(0.452)	(1.482)	(0.638)	(1.02)
PDY (T-1)		-0.012				-0.005		
		(0.007)				(0.02)		
ΔPUBINVY			-0.17				-0.403	
			(0.422)				(0.398)	
ΔSOCIALY				0.642				0.583
				(0.423)				(0.482)
Constant	0.55	2.032**	-0.172	-0.6	0.008	0.126	-0.418	-0.9
	(0.587)	(1.035)	(0.52)	(0.66)	(0.516)	(1.111)	(0.515)	(0.835)
N	74	63	58	30	66	52	58	29
R^2	0.629	0.642	0.748	0.822	0.686	0.677	0.731	0.846
SE (error term)	1.508	1.445	1.373	1.201	1.438	1.463	1.434	1.197

Source: IEO staff estimates.
Notes: Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

54. The results presented in Table 8 for PRGT programs suggest that PRGT programs have on average assumed smaller fiscal multipliers than GRA programs (columns (1)-(5)), which is consistent with the findings in the literature that fiscal multipliers are generally smaller in emerging and LICs than in advanced economies.[15] As in GRA programs, the estimated coefficients of ΔGREVY and ΔGEXPY in columns (6)-(10) indicate that PRGT programs also have on average assumed revenue multipliers to be smaller than expenditure multipliers in program design although the difference between them is less marked than that in GRA programs. Finally, PRGT programs are like GRA programs in that no statistically significant tradeoff is detected between CA adjustment and growth projections.

55. As to external linkages in the macroeconomic framework, growth projections are rather tightly linked to external demand conditions captured by trading partners' growth (TP_GR) in GRA programs. In PRGT programs, linkage to external demand condition is significantly weaker than in GRA programs, but initial growth seems to have affected program growth projections more prominently. The impact on growth of public investment and social spending is found to be statistically insignificant in both GRA and PRGT programs except for one case in PRGT programs.

56. All in all, the regression results in Tables 7 and 8 suggest that the macroeconomic framework envisaged in initial program design has on average incorporated short-run tradeoffs between fiscal adjustment and growth with assumed fiscal multipliers being on the order of 0.3-0.5 in GRA programs and about 0.2 in PRGT programs. Moreover, program assumptions on revenue and expenditure multipliers in initial program design appears broadly consistent with existing evidence in the literature. Not surprisingly, no tight relationship is detected for CA adjustment and growth program projections in both types of programs.

Estimation Results for Program Updates

57. The same regression analysis is undertaken to assess the macroeconomic frameworks used in updating program projections. As discussed in Section V, program projections are revised over the program period in response to interim macroeconomic outcomes. As such, the regression analysis based on updated (one-period-ahead) program projections shed light on how the macroeconomic framework has evolved over the program period to incorporate new information from interim adjustment and growth outcomes. For consistency, projection data used for this estimation exercise consist of initial projections for the first year (T) of the program (which are one-year-ahead projections) and one-year-ahead projections for the rest of the program period.[16]

[15] See Batini and others (2014), IMF (2017), and Gupta (2021) as well as the references therein for discussions on the size of fiscal multipliers of various income groups. IMF (2017) finds that estimated fiscal multipliers in Sub-Saharan Africa tend to be smaller than those typically identified in advanced or emerging market economies. Batini and others (2014) list a range of estimated fiscal multipliers from the scarce empirical literature on EMEs and LICs, which indicates fiscal multipliers are generally smaller in emerging and LICs than in advanced economies.

[16] Note that the sample size differs from the one used for Tables 7 and 8 based on initial program projections because actual program duration does not always coincide with program duration envisaged in initial program design.

Table 8. Results of PRGT Growth Regressions: Initial Program Projections

	PRGT							
	(1) Growth	(2) Growth	(3) Growth	(4) Growth	(5) Growth	(6) Growth	(7) Growth	(8) Growth
ΔPBY	-0.175**	-0.175	-0.077	-0.223**				
	(0.071)	(0.11)	(0.112)	(0.104)				
ΔGREVY					-0.162**	-0.17	0.025	-0.141**
					(0.067)	(0.151)	(0.104)	(0.065)
ΔGEXPY					0.165*	0.186	0.149	0.201*
					(0.091)	(0.234)	(0.147)	(0.116)
ΔCABY	0.051	0.048	0.064	0.073	0.058	0.051	0.057	0.098
	(0.044)	(0.049)	(0.055)	(0.052)	(0.055)	(0.058)	(0.055)	(0.07)
TP_GR	0.123	-0.049	0.103	0.23**	0.14*	-0.063	0.132	0.217*
	(0.079)	(0.108)	(0.093)	(0.111)	(0.083)	(0.155)	(0.092)	(0.121)
ΔTOT	-0.027*	-0.016	-0.056**	-0.03	-0.037*	-0.022	-0.059**	-0.016
	(0.016)	(0.023)	(0.024)	(0.029)	(0.021)	(0.031)	(0.025)	(0.035)
Growth (T-1)	0.197***	0.200***	0.201***	0.093	0.157**	0.165**	0.2***	0.077
	(0.058)	(0.06)	(0.069)	(0.096)	(0.067)	(0.066)	(0.071)	(0.094)
EDY (T-1)	-0.003	-0.007	-0.009	-0.010	-0.006	-0.01	-0.008	-0.013*
	(0.006)	(0.008)	(0.006)	(0.007)	(0.007)	(0.011)	(0.006)	(0.007)
GEI (T-1)	0.501	0.557	-0.162	0.838	0.256	-0.038	0	0.514
	(0.345)	(0.427)	(0.41)	(0.634)	(0.42)	(0.665)	(0.412)	(0.585)
Small states	-1.133***	-0.985**	-1.054	-1.02*	-1.37***	-1.255**	-0.683	-1.024*
	(0.383)	(0.414)	(0.793)	(0.576)	(0.431)	(0.494)	(0.784)	(0.567)
Fragile states	0.039	0.206	-0.169	0.235	0.141	0.149	-0.007	0.321
	(0.403)	(0.454)	(0.44)	(0.506)	(0.39)	(0.462)	(0.416)	(0.438)
PDY (T-1)		0.003				0.001		
		(0.007)				(0.008)		
ΔPUBINVY			0.316*				0.234	
			(0.168)				(0.182)	
ΔSOCIALY				0.189				0.094
				(0.16)				(0.212)
Constant	4.862***	5.587***	4.805***	5.478***	4.965***	5.673***	4.75***	5.517***
	(0.685)	(0.762)	(0.751)	(1.095)	(0.739)	(0.877)	(0.777)	(1.094)
N	189	163	135	100	159	130	138	90
R^2	0.21	0.18	0.337	0.276	0.21	0.177	0.349	0.287
SE (error term)	2.073	2.181	1.875	1.642	2.224	2.398	1.857	1.669

Source: IEO staff estimates.
Notes: Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

58. Full regression results for updated program projections are reported in Tables AVI.2 and AVI.3 in Appendix VI. The regression results paint somewhat different picture about fiscal multipliers from what can be inferred from Tables 7 and 8 about initial program design. Specifically:

- First, updated fiscal multipliers for GRA programs are on the order of 0.24–0.35, which are in general smaller than those assumed in initial program design. In contrast, updated fiscal multipliers for PRGT programs are on the order of 0.1-0.3 (albeit rarely significant) and on average modestly larger than assumed in initial program design.

- Second, updated revenue multipliers are on average larger, and not smaller, than expenditure multipliers in both GRA and PRGT programs. Moreover, none of the updated expenditure multipliers are statistically significant. These results are at odds with existing evidence in the literature.

59. In sum, the regression results for initial and updated projections suggest that the macroeconomic frameworks used in program design have embodied a short-run tradeoff between fiscal adjustment and growth broadly consistent with the literature. Program assumptions on fiscal multipliers seem to have been recalibrated as new information becomes available from interim adjustment and growth outcomes. According to the regression results, initial assumptions on fiscal multipliers seem to have ranged around 0.35–0.5 in GRA programs and 0.17–0.22 in PRGT programs with revenue multipliers being in general smaller than expenditure multipliers. Assumptions used in program update seem to have settled on smaller multipliers than initially assumed in GRA programs but modestly higher multipliers in PRGT programs (although rarely significant in PRGT programs), as well as revenue multipliers larger, and not smaller, than expenditure multipliers.

VII. BENCHMARKING GROWTH OUTCOMES

60. This section develops a benchmark against which to compare growth outcomes under IMF-supported programs that seeks to take account of exogenous shifts in the external environment. IMF-supported programs in the evaluation sample were approved and completed at different times, against different situations for the global economy; countries also experienced different terms of trade and external demand shocks depending on their economic structure and regional context; moreover, program countries in the sample differ widely in historical growth trends. These differences pose an empirical challenge in making consistent cross-country comparison of growth outcomes over programs which span a few years at most. In this respect, growth outcomes need to be measured relative to a common benchmark that is comparable across time and countries. This section discusses how the benchmarks for actual and trend GDP growth used in the cross-country analysis were estimated and the estimation results.

61. The growth benchmarks considered are intended to capture the variation in actual or potential GDP growth explained by external factors alone, including factors affecting demand for a country's exports and availability of external market financing. As such, the difference between actual and benchmark growth rates can be interpreted as primarily reflecting the influence of domestic factors such as domestic policies and supply shocks. It should be noted that the benchmark is not a counterfactual (e.g., growth outcome that would have prevailed for different policies or with no Fund engagement). Rather, the benchmark should be considered as an ex-post estimated component of growth which is exogenous to IMF-supported programs and beyond the control of country authorities or the IMF.

62. Specifically, the benchmark for actual growth is estimated by using a panel OLS regression with country fixed effects for a large sample of 174 countries (with and without IMF programs) over the period of 1990–2019. The explanatory variables include the percentage change in the country's terms of trade (ΔTOT), trading partners' growth rate (TP_GR), regional growth rate (RG_GR), and the US policy interest rate (US_Rate).[17] Country fixed effects and a dummy for the period of 2008–19 are included to reflect different trend growth across countries and to capture the lasting effect on growth of the global financial crisis (GFC), respectively. It should be noted that no control variables are included to distinguish between program and non-program countries or periods in order to make the deviation of actual growth from the benchmark comparable across time and countries on a consistent footing.

63. The benchmark for potential growth, which is used in the cross-country analysis of post-program growth in Section XI, is estimated by using the same specification as used for the benchmark for actual growth. To be specific, the potential growth rate is obtained as the rate of growth of real potential GDP which is estimated by applying the HP filter to real GDP for each country. The explanatory variables included in the benchmark growth regression for potential growth are the trend component of the same variables used in the estimation of the benchmark for actual growth.

64. The panel regression results for actual and potential growth are shown in Table 9.[18] The results for actual growth are displayed in columns (1)-(3) and potential growth in column (4). The estimated coefficients are of the expected sign and mostly highly statistically significant, except for the US policy rate in column (2) and the post-GFC dummy in column (3). Despite the parsimonious specification, the regressions for actual growth account for 30-50 percent of total variation in the actual growth rates. The estimated coefficients for potential growth in column (4) are also of the expected sign for all explanatory variables, and statistically significant except for regional growth. The estimated regression explains about 60 percent of the sample variation in the potential growth rates.

65. The benchmarks for actual and potential growth are constructed by using the baseline specifications presented in columns (3) and (4) of Table 9, respectively. In what follows, AGBD and PTGBD stand for the deviation of actual and potential growth rates from their benchmarks, respectively. In conceptual terms, AGBD and PTGBD reflect primarily the influence of domestic policies and shocks and are suited for more consistent cross-country comparison of growth performance given that country fixed effects are purged. The remainder of this section focuses on cross-country comparison of AGBD during the program period. PTGBD is used in the regression analysis of post-program potential growth discussed in Section XI.

[17] A small open economy assumption is invoked to treat ΔTOT as an exogenous variable and validate the use of OLS-FE regression.

[18] The regressions are estimated for a sample which removed large outliers (top and bottom 1 percent) in the data for the growth rate and the terms of trade.

Table 9. Panel Growth Regression Results (OLS-FE)

	Actual Growth[1]			Potential Growth[1, 2]
	(1) 1990-2007	(2) 2008-2019	(3) 1990-2019	(4) 1990-2019
ΔTOT	0.013***	0.025***	0.017***	0.088***
	(0.005)	(0.007)	(0.005)	(0.014)
TP_GR	0.394***	0.452***	0.443***	1.164***
	(0.060)	(0.023)	(0.047)	(0.156)
RG_GR	0.446***	0.353***	0.410***	0.031
	(0.042)	(0.022)	(0.044)	(0.044)
US_Rate	-0.050**	-0.007	-0.051**	-0.280***
	(0.024)	(0.033)	(0.026)	(0.042)
Post-GFC dummy			0.012	-0.161***
			(0.139)	(0.050)
No. of observations	2636	1984	4620	4618
No. of countries	167	174	174	174
Overall R^2	0.308	0.477	0.300	0.587
SE of regression	3.455	2.481	3.238	1.556
Country FE	Yes	Yes	Yes	Yes
Time FE	No	No	No	No

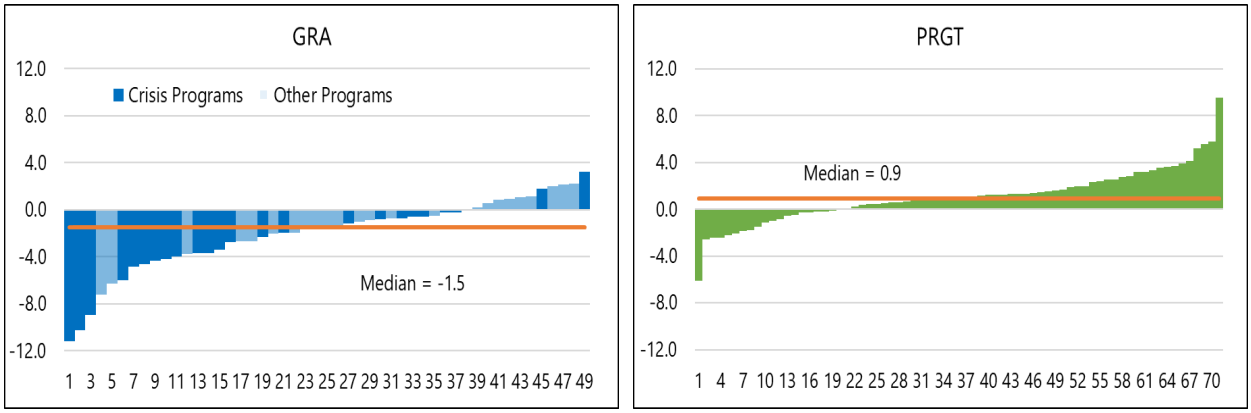
Source: IEO staff estimates.
[1] Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.
[2] Explanatory variables represent trend growth or component except for the post-GFC dummy.

66. AGBD provides a useful picture of growth outcomes of IMF-supported programs on a consistent basis. To this end, the annual average of AGBD over the program period is constructed for 120 programs with actual program duration of one year or longer. Annual averages of AGBD range from -11.2 percent for Ukraine (2008 SBA) to 9.5 percent for Afghanistan (2011 PRGT), with the cross-country median of 0.2 percent in the full sample (Figure 16). While the full sample is split relatively evenly between positive AGBDs (64 programs) and negative AGBDs (56 programs), the GRA sample is populated largely by negative AGBDs while the opposite holds for the PRGT sample, reflecting in part that GRA programs tend to focus more on stabilization in the face of acute BOP pressure than PRGT programs. As a result, the sample median diverges significantly between GRA (-1.5 percent) and PRGT programs (0.9 percent).

67. Statistically significant AGBDs, which account for 20 percent of the total, are even more unevenly distributed between the GRA and PRGT samples.[19] Specifically, positive and significant growth deviations are entirely from the PRGT sample while almost all negative and significant deviations are from the GRA sample (Table 10). Within the GRA sample, crisis programs dominate other programs as a source of negative and significant growth deviations—11 out of 13 negative and significant deviations in the GRA sample are associated with crisis programs. Overall, AGBDs are negative and statistically significant in less than 12 percent of the full sample.

[19] AGBDs are assumed to be normally distributed with zero mean given that they are the residuals of the estimated panel regression. Let AGBD(n) denotes the annual average of AGBD of a program with duration of n years. Assuming no serial correlation, it is straightforward to show that z(n) = AGBD(n)/SE(n) follows the standard normal distribution where $\mathrm{SE}(\mathrm{n}) = \mathrm{SE}/\sqrt{\mathrm{n}}$ and SE is the standard error of the baseline panel regression in column (3) of Table 10. Statistical significance of AGBD(n) is based on the value of z(n).

Figure 16. Distribution of AGBD Across Programs
(In percent; annual averages)



Source: IEO staff estimates.

Table 10. Distribution of AGBD by Program Type

Program Type	Positive	Negative	Total
GRA	12 (0)	37 (13)	49 (13)
Crisis	2 (0)	20 (11)	22 (11)
Other	10 (0)	17 (2)	27 (2)
PRGT	52 (10)	19 (1)	71 (11)
Total	64 (10)	56 (14)	120 (24)

Source: IEO staff estimates.
Note: The numbers in parentheses are the number of programs for which the annual average of AGBD is statistically significant different from zero at 10 percent or higher.

68. In sum, these results suggest that IMF-supported programs under evaluation have generally fared relatively well in terms of growth once the influence of exogenous external factors and country fixed effects is controlled for. It appears that PRGT programs have on average fared significantly better than GRA programs with AGBD being higher on average by 3.1 percentage points, which is not surprising given that GRA programs tended to focus more on stabilization in the face of acute BOP pressure than PRGT programs. This latter finding suggests that the differences in external conditions or historical trend growth would explain only one quarter of the total difference of 4.3 percentage points in average growth outcomes between GRA and PRGT programs (see Table AVI.1 in Appendix VI). The remaining three quarters could be related at least in part to different adjustment needs which were usually larger in GRA programs than in PRGT programs as can be seen from Figures 8 and 10 in Section III.

VIII. MODELING GROWTH IN PROGRAM OUTCOMES

69. The main objective of this section is to estimate the short-run relationships between adjustment and growth outcomes in GRA and PRGT programs, which can be compared with the estimation results for program design in Section VI to assess the realism of the macroeconomic frameworks of the programs in the evaluation sample.

70. The regression analysis in this section differs in two respects from the regression analysis for program design in Section VI. First, estimation is conducted by using cross-section data for annual averages over the program period. The focus on cross-section regressions is motivated by the desire to obtain sharper estimates of fiscal multipliers in light of the fact that in practice the growth impact of fiscal adjustments tends to materialize over a period longer than a year.[20] For instance, fiscal adjustment undertaken at the beginning of year t may affect growth not only in year t but also year t+1 and even beyond. If fiscal adjustment is instead undertaken at the end of year t, it would little affect growth in year t, and most of its short-run growth impact would likely materialize in year t+1 and afterwards. Second, growth outcomes are measured by AGBD—i.e., the deviation of actual growth from the common benchmark discussed in Section VII—rather than actual growth because the former is more comparable across programs and time than actual growth.[21]

71. The regression analysis for program outcomes uses essentially the same specifications considered for the analysis of program design as shown in Tables 7 and 8, except that two external variables—trading partners' growth and the terms of trade—are omitted from the regressions since their influence on growth is already taken care of by using AGBD as the dependent variable.

Data and Estimation Results

72. Program outcome data used for the regression analysis are all taken from the January WEO vintage of 2020.

73. Regressions for growth outcomes are estimated by using essentially the same specifications with those considered for program projection regressions discussed in Section V. Unlike in the regressions based on program projection data, program outcome regressions are subject to simultaneity bias. For instance, innovations in fiscal adjustment (ΔPBY) are likely to be

[20] The use of cross-section data—i.e., program period averages—would help to average out the cyclical component in the primary balance measured on the annual frequency and bring the resulting average primary balance closer to the structural primary balance. Although not reported, the panel growth regressions for program outcomes are estimated imprecisely producing no meaningful results for fiscal multipliers as well as other coefficients.

[21] As the analysis relies on cross-section regressions while programs in the estimation sample were arranged at different times, it is not possible to control for different historical trend growth across countries if actual growth is used as the dependent variable. Use of AGBD is relatively free of this problem since country fixed effects are included in the panel regression in Table 9, making it more comparable across programs.

positively correlated with innovations in growth if the latter lead to higher government revenue than otherwise. Such positive correlation arising from reverse causality (from growth to fiscal adjustment) could bias the coefficient of ΔPBY upward, in which case actual fiscal multipliers would be underestimated. To address the simultaneity bias, IV regressions are used in which explanatory variables considered to be endogenous to innovations in growth are instrumented. Specifically, ΔPBY, ΔGREVY, ΔGEXPY, and ΔCABY are instrumented by one-period-ahead staff projections while ΔREER is instrumented by its own lagged values.

74. Table 11 shows the results of cross-section growth regressions for the full, GRA and PRGT samples. Most explanatory variables are of the expected sign and, in many cases, statistically significant.[22] The estimated coefficients of ΔPBY imply actual fiscal multipliers on the order of 0.35-0.45 in GRA programs and 0.6 in PRGT programs (columns (3)-(5) and (11)). This result is somewhat surprising and at odd with existing evidence that fiscal multipliers are found to be smaller in LICs eligible for the PRGT. Moreover, the results for revenue and expenditure multipliers indicate that revenue-based adjustment were significantly more contractionary than expenditure-based adjustment in the program context (columns (7)-(10) and (12)), which is again at odd with existing evidence in the literature.[23]

75. The relationship between CA adjustment and growth outcomes is negative and statistically significant in PRGT programs but not in the GRA sample. This result is broadly consistent with the composition of actual CA adjustment between exports and imports shown in Figure 10—i.e., actual CA adjustment relied on import compression relatively more heavily in PRGT programs than in GRA programs. The estimated coefficients of ΔREER indicate that real depreciation (ΔREER <0) has helped to boost growth, although the effect is found to be statistically significant for PRGT programs but not for GRA programs (except where a crisis program dummy is included). To be more specific, the results for PRGT programs suggest that a 10 percent depreciation in real effective terms would boost the annual growth on average by about 1.1-1.7 percentage points. For GRA programs, the range for estimated exchange rate impact on growth is much wider (0.4-3.0 percentage points for a 10 percent real depreciation) although generally not significant.

[22] In GRA regressions, not all explanatory variables are included at once because of limited number of observations. Instead, some explanatory variables are included individually across specifications.

[23] This result may not hold for a larger and more balanced sample that includes both program and non-program countries and periods. It may also reflect some aspects of program design specific to the sample. In some programs, for instance, tax increases were achieved primarily by new taxes or an increase in the tax rate, rather than a broadening of the tax base and a strengthening of tax administration. In other programs, expenditure cuts were focused on reducing inefficient current expenditure or cutting capital spending on import-intensive projects while protecting social spending.

Table 11. Results of Growth Regressions (IV): Program Outcomes

	FULL		GRA								PRGT	
	(1) AGBD	(2) AGBD	(3) AGBD	(4) AGBD	(5) AGBD	(6) AGBD	(7) AGBD	(8) AGBD	(9) AGBD	(10) AGBD	(11) AGBD	(12) AGBD
ΔPBY	-0.48***		-0.448***	-0.16	-0.403**	-0.345**					-0.61*	
	(0.122)		(0.154)	(0.213)	(0.177)	(0.144)					(0.341)	
ΔGREVY		-0.906***					-1.48***	0.864	1.488	-1.123**		-1.021**
		(0.225)					(0.481)	(0.692)	(1.343)	(0.473)		(0.513)
ΔGEXPY		0.408***					0.34**	0.251	0.634**	0.262		0.052
		(0.143)					(0.159)	(0.248)	(0.322)	(0.162)		(0.487)
ΔCABY	-0.123	-0.192	0.285	0.628***	0.507*	0.373	0.103	0.921***	0.902*	0.202	-0.449**	-0.665***
	(0.113)	(0.12)	(0.296)	(0.222)	(0.266)	(0.295)	(0.296)	(0.306)	(0.548)	(0.309)	(0.188)	(0.235)
ΔREER	-0.096**	-0.104***	-0.135	-0.047	-0.118	-0.086	-0.065	-0.116	-0.276	-0.038	-0.111**	-0.166***
	(0.038)	(0.039)	(0.124)	(0.114)	(0.104)	(0.104)	(0.123)	(0.144)	(0.184)	(0.112)	(0.048)	(0.06)
YGAP (T-1)	-0.38***	-0.386***	-0.545***	-0.493***	-0.468***	-0.485***	-0.429***	-0.622***	-0.706***	-0.414***	-0.357***	-0.342***
	(0.055)	(0.054)	(0.098)	(0.087)	(0.101)	(0.095)	(0.12)	(0.128)	(0.22)	(0.113)	(0.064)	(0.105)
PDY (T-1)	0.523	0.59	-0.802	-0.563	-0.484	-0.247	0.117	-1.436	-1.941	0.271	1.032	0.83
	(0.471)	(0.455)	(0.611)	(0.656)	(0.655)	(0.519)	(0.812)	(0.936)	(1.351)	(0.682)	(0.809)	(1.013)
EDY (T-1)	-0.548*	-0.504*	-1.741***	-1.371***	-0.819**	-1.068***	-1.316**	-1.655***	-1.208*	-0.977***	-0.253	0.011
	(0.296)	(0.295)	(0.57)	(0.409)	(0.377)	(0.277)	(0.637)	(0.408)	(0.706)	(0.274)	(0.6)	(0.669)
GEI	-0.062	-0.292	1.735*				1.125				0.849	0.439
	(0.419)	(0.445)	(0.973)				(1.109)				(0.636)	(0.732)
PRGT programs	-0.935*	-1.413**										
	(0.527)	(0.589)										
Small states	1.047*	0.814		3.399**				4.486*			3.199*	2.891
	(0.589)	(0.594)		(1.69)				(2.359)			(1.664)	(1.979)
Fragile states	1.53**	1.75**									0.091	0.169
	(0.739)	(0.825)									(0.969)	(1.06)
Crisis programs	-0.008	-0.263			-0.795				-0.663			
	(0.588)	(0.609)			(0.894)				(1.265)			
Off-track programs	0.413	0.524				-1.299*				-1.197*	-0.72	-2.461**
	(0.68)	(0.655)				(0.672)				(0.647)	(0.686)	(1.233)
Constant	-0.35	-0.557	9.721**	6.416**	4.577*	4.86**	4.439	10.817***	11.735*	2.585	-1.582	-1.885
	(1.819)	(1.78)	(4.115)	(2.919)	(2.61)	(2.3)	(4.915)	(3.943)	(6.421)	(2.898)	(1.852)	(2.301)
N	88	87	34	34	34	34	33	33	33	33	54	54
R-sd overall	0.711	0.719	0.733	0.744	0.7	0.739	0.763	0.648	0.421	0.778	0.534	0.437
SE (error term)	1.586	1.589	1.606	1.574	1.701	1.587	1.57	1.95	2.655	1.519	1.708	2.031

Source: IEO staff estimates.
Note: Robust standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01. Parsimonious specifications are used for GRA programs due to small sample.

76. Regarding initial conditions, the initial output gap is found to be an important and statistically significant determinant of program growth performance, indicating that growth recovery during the program has on average been stronger the deeper was the initial recession (YGAP < 0). Higher initial indebtedness, both public and external (i.e., PDY(T-1) and EDY(T-1)), has affected growth negatively during the program, particularly in GRA-supported programs.

Comparison Between Program Design and Outcomes

77. Strictly speaking, the regression results for program outcomes in Table 11 are not fully comparable with the regression results for program projections in Tables 7-8 (initial projections) and Tables AVI.2-AVI.3 in Appendix VI (updated projections), because of the differences in the

measurement of the growth variable and data frequency. Notwithstanding these differences, some comparison can still be made for fiscal multipliers.[24]

- First, initial program assumptions on overall fiscal multipliers appear broadly consistent with actual overall multipliers in GRA programs but appear to have understated actual multipliers in PRGT programs.
- Second, updated program assumptions on overall fiscal multipliers seem on average less realistic than initial assumptions in GRA programs but more realistic in PRGT programs (albeit still somewhat smaller than actual multipliers).
- Third, in both GRA and PRGT programs, program assumptions appear to have understated actual revenue multipliers although updated assumptions seem more realistic than initial program assumptions.
- Fourth, updated program assumptions on expenditure multipliers have deviated further from actual multipliers than initial program assumptions, particularly in GRA programs.

78. Taken together, these findings suggest that the initial design of the macroeconomic framework was more realistic about fiscal multipliers than when updated in GRA programs, while the opposite seems to be the case in PRGT programs.

79. For the relationship between CA adjustment and growth, the results for program projections and outcomes are broadly consistent in GRA programs in the sense that the relationship is statistically insignificant in most cases in both projection and outcome data. In PRGT programs, however, the actual relationship is negative and statistically significant while program projections have envisaged a positive (albeit statistically insignificant) relationship if any.

IX. Macroeconomic Modeling and Growth Optimism

80. It has long been recognized that the IMF's growth forecasts are subject to optimism bias in both surveillance and program contexts. The 2018 ROC confirmed optimism bias in growth projections, following the approach used by Blanchard and Leigh (2013).

81. This section uses a cross-country analysis of growth forecast errors to assess how important macroeconomic modeling errors were as a source of growth optimism in the program context. Optimism bias in growth projections could arise from unrealistic program assumptions on the behavioral relationship between adjustment and growth. This analysis complements the previous analysis of the macroeconomic framework in program design and outcomes discussed respectively in Sections VI and VIII.

[24] The difference in the measurement of growth is less of an issue for comparability given that two major external variables–trading partners' growth (TP_GR) and the percentage change in the terms of trade (ΔTOT)—are controlled for in program projection regressions reported in Tables 7-8 and Tables AVI.2-AVI.3.

Data and Estimation Results

82. Drawing on the approach used by the 2018 ROC, Blanchard-Leigh type growth forecast error regressions are estimated for 75 program countries in the evaluation sample over the period of 2009–19. The use of the full sample period of 2009–19 is motivated to allow comparisons between program and non-program periods. In the regression analysis, growth forecast errors are defined as projection minus actual so that positive (negative) forecast errors indicate growth optimism (pessimism).

83. Unlike in Blanchard and Leigh (2013) and the 2018 ROC, growth forecast error regressions are estimated for one-period-ahead forecast errors, rather than two-period-ahead forecast errors for two reasons. First, one-period-ahead forecast errors better capture the reality given that the macroeconomic framework of the program is updated over the program period incorporating new information available from interim program outcomes. Second, our objective is not to estimate fiscal multipliers with high precision but to assess the contribution of macroeconomic modeling errors to sample variation in growth forecast errors.

84. To construct forecast errors, the actual data are all taken from the 2020 January WEO vintage while the projection data are taken from the April WEO vintage of each year in the sample period.[25] An exception is for program projections for the first year (T) of the program which are taken from the first WEO vintage published after program approval.

85. Table 12 displays the results of growth forecast error regressions. The first two explanatory variables, FE_TPGR and FE_ΔTOT represent forecast errors in trading partners' growth and the terms of trade, respectively. The rest of the explanatory variables are projections on fiscal and CA adjustments (ΔPBY and ΔCABY) and their interaction terms with various program dummies—PROG for the program period, GRA and PRGT for program type, and Crisis for crisis programs as defined in Table 1. LGFA is a dummy that takes 1 for fiscal adjustment projections larger than 2.0 percent of GDP and 0 otherwise. LGCA is a dummy similarly defined for large CA adjustment projections above 3.22 percent of GDP.[26] The regressions include country and vintage fixed effects which are not reported.

[25] To be specific, projections for growth in year t are taken from the April WEO vintage of year t. Use of the April WEO vintage for projection data is consistent with the approach adopted in Blanchard and Leigh (2013). Growth forecast error regressions are estimated less precisely if projection data are instead taken from the October vintage.

[26] The thresholds for LGFA and LGCA are the values which maximize R^2 and found by using a grid search.

Table 12. Results of Growth Forecast Error Regressions

	(1) FE_GR	(2) FE_GR	(3) FE_GR	(4) FE_GR	(5) FE_GR
FE_TPGR	-0.02 (0.101)	-0.026 (0.104)	-0.02 (0.101)	-0.013 (0.1)	-0.042 (0.1)
FE_ΔTOT	0.029*** (0.008)	0.029*** (0.008)	0.029*** (0.008)	0.028*** (0.008)	0.029*** (0.008)
ΔPBY	0.025 (0.035)	0.111** (0.048)	-0.014 (0.05)	-0.015 (0.05)	-0.022 (0.049)
ΔCABY	-0.027 (0.029)	-0.067* (0.038)	-0.018 (0.045)	-0.02 (0.044)	-0.015 (0.044)
ΔPBY*LGFA		-0.161** (0.077)			
ΔCABY*LGCA		0.09 (0.065)			
ΔPBY*PROG			0.094 (0.07)		
ΔCABY*PROG			-0.009 (0.053)		
ΔPBY*GRA				0.289** (0.141)	
ΔPBY*PRGT				0.042 (0.071)	0.049 (0.07)
ΔCABY*GRA				0.011 (0.063)	
ΔCABY*PRGT				-0.011 (0.06)	-0.018 (0.06)
ΔPBY*GRA_Crisis					0.191 (0.174)
ΔPBY*GRA_Other					0.49** (0.242)
ΔCABY*GRA_Crisis					0.002 (0.089)
ΔCABY*GRA_Other					-0.116 (0.113)
PROG			-0.042 (0.21)		
GRA				-0.051 (0.375)	
PRGT				-0.178 (0.259)	-0.198 (0.258)
GRA_Crisis					0.868 (0.668)
GRA_Other					-1.069** (0.464)
Constant	0.035 (1.092)	0.019 (1.081)	0.054 (1.089)	0.117 (1.098)	0.045 (1.097)
N	703	703	703	703	703
R-sd overall	0.259	0.268	0.262	0.269	0.284
SE (error term)	2.089	2.08	2.09	2.086	2.07

Source: IEO staff estimates.
Note: Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

86. Overall, the regressions results explain about one-quarter of total sample variation in growth forecast errors, which is broadly in line with the findings of the 2018 ROC. The result in column (1) indicates a significant impact of terms of trade shocks but no statistically significant contribution from fiscal or CA projections to growth forecast errors. If nonlinearity is allowed as shown in column (2) by introducing the "large adjustment" items, however, the coefficient on fiscal adjustment is significant, and suggests that large fiscal adjustments were associated with smaller optimism bias than small or moderate fiscal adjustments. While this latter finding looks at odds with evidence found by the 2018 ROC and other related studies,[27] it is appealing in the sense that it suggests that country teams may have factored in confidence effects in program projections which large fiscal adjustments can entail and help to offset in part income effects.

87. The regression results in columns (3)-(5) enable comparison between program and non-program periods. The result in column (3) indicates that growth forecast errors are again significantly affected by forecast errors in the terms of trade, but little by fiscal or CA projections in both program and non-program periods. In addition, the coefficient of the program dummy (PROG) is statistically insignificant.

88. When program periods are further disaggregated into GRA and PRGT program periods, however, the regression result shows that fiscal projections in GRA programs are a statistically significant source of growth optimism while fiscal projections in PRGT programs or CA projections are not (column (4)). This result is broadly consistent with the findings in Sections III and VIII that optimism bias in updated projections is on average larger for GRA programs than for PRGT programs (see Figure 6), while updated program assumptions on fiscal multipliers were on average less realistic (and smaller) than initial program assumptions in GRA programs but less so in PRGT programs (see Tables AVI.2-AVI.3 in Appendix VI).[28] Finally, the result in column (5) shows that fiscal projections in GRA programs other than crisis programs, are an important source of optimism bias within GRA programs. This finding is broadly consistent with the result in column (2) which indicates on average a smaller optimism bias for large fiscal adjustments (which are more likely in crisis programs than in other GRA programs).

89. Further investigation reveals that fiscal forecast errors (defined as projection minus actual) are positively and significantly correlated with fiscal projections in GRA crisis and PRGT programs, indicating that in these programs fiscal projections tended to overpredict fiscal outturns by more the larger is programmed fiscal adjustment. This finding implies that the estimated coefficients of ΔPBY in crisis and PRGT programs in growth forecast error regressions are likely to be biased downwards and hence actual fiscal multipliers in these programs could be

[27] For instance, Ismail and others (2020) find for a large panel sample of 170 countries for the period of 2003–17 that larger fiscal adjustment is associated with higher growth optimism in GRA programs.

[28] Growth forecast errors for any given year in Table 12 are based on projection data taken from the April vintage of the WEO of the same year following Blanchard and Leigh (2013), while the growth regressions in Tables AVI.2-AVI.3 in Appendix VI are based on projection data taken from the October vintage of the previous year. Nonetheless, growth projections from the two vintages are little different.

higher than implied by the results in Table 12. In contrast, no significant correlation is found for GRA programs other than crisis programs. This latter finding suggests that the statistically significant effect of fiscal projections on growth forecast errors is primarily related to unrealistic program assumptions on fiscal multipliers rather than slippages in the implementation of fiscal adjustment (which are reflected in fiscal forecast errors in the data).

Decomposing Growth Optimism

90. The findings on growth optimism from the regression analysis are further confirmed by the results of variance decomposition for growth forecast error for various subsamples. Table 13 reports the variance ratios constructed by using the specification in column (5) of Table 12. Variance ratios do not necessarily add up across columns and could exceed 100 percent because of non-zero covariances.[29] Panels A and B show the results of variance decomposition undertaken, respectively, for explained and total sample variation in growth forecast errors after country and vintage fixed effects.

Table 13. Growth Forecast Error Variance Decomposition

A. Explained Variation
(In percent)

Period/Program	(1) ΔPBY	(2) ΔCABY	(3) ΔPBY + ΔCABY	(4) FE_TPGR + FE_DTOT
Non-program	4.6	3.3	11.3	92.4
Program	59.5	15.5	63.3	37.9
PRGT	4.2	11.0	11.0	91.4
GRA	88.6	18.1	94.8	9.4
Crisis	59.5	1.7	66.7	25.0
Other	96.5	22.7	103.2	4.7

B. Total Variation
(In percent)

Period/Program	(1) ΔPBY	(2) ΔCABY	(3) ΔPBY + ΔCABY	(4) FE_TPGR + FE_DTOT
Non-program	0.1	0.1	0.3	2.5
Program	4.7	1.2	5.0	3.0
PRGT	0.2	0.5	0.5	4.3
GRA	10.0	2.0	10.7	1.1
Crisis	2.2	0.1	2.5	0.9
Other	26.4	6.2	28.3	1.3

Source: IEO staff estimates.
Note: The numbers are variance ratios and do not add up across columns due to non-zero covariances. Variance ratios are calculated as the ratio of variance of each variable to explained or total sample variation in growth forecast errors after country and vintage fixed effects.

[29] For example, consider a regression equation given by $y = b_0 + b_1x_1 + b_2x_2 + \varepsilon$ where y is growth forecast errors after country and vintage fixed effects and ε is a residual. Variance ratio for x_1 relative to explained variation is calculated as $VR = Var(b_1x_1)/Var(b_1x_1 + b_2x_2)$. Similarly, variance ratio for x_1 relative to total variation is calculated as $VR = Var(b_1x_1)/Var(y)$.

91. The variance decomposition results in Panel A show that the combined contribution of projected fiscal and CA adjustments to explained variation in growth forecast errors is 95 percent in GRA programs, which dwarfs the same combined contribution of 11 percent in PRGT programs and outside programs (column (3)) where growth forecast errors are largely explained by unforeseen shocks to the terms of trade and trading partners' growth. In addition, the contribution of fiscal projections dominates the contribution of CA projections in program periods (columns (1) and (2)).

92. Panel B shows the variance decomposition results for total sample variation. Variance ratios in Panel B are significantly smaller than corresponding variance ratios in Panel A, indicating that unexplained sample variation is large even after country and vintage fixed effects.[30] Nonetheless, the contribution of fiscal projections to total sample variation remains significant at 28 percent in GRA programs other than crisis programs while the results for crisis programs are in line with the results for PRGT programs or non-programs.

93. Taken together, the variance decomposition results reinforce the finding in Section VIII that the macroeconomic framework used to underpin program growth projections in GRA programs is on average less realistic—particularly with respect to the growth impact of fiscal adjustment—than the framework used in PRGT programs and outside the program context. In GRA programs other than crisis programs, fiscal projections were on average accurate (see Figure 8) and fiscal forecast errors are broadly uncorrelated with fiscal projections. This in turn implies that the significant contribution of fiscal projections to variation in growth forecast errors found in these programs is primarily a result of unrealistic program assumptions on fiscal multipliers, rather than slippages in fiscal policy implementation.

X. Short-Run Growth Impact of IMF-Supported Programs

94. This section presents an empirical analysis of the short-run (i.e., during the program period) growth impact of IMF-supported programs. It is recognized that growth outcomes during IMF-supported programs should be viewed against the extraordinary circumstances faced by program countries and substantial adjustment needed to restore external viability. Thus, stabilization programs would necessarily involve restraints on aggregate demand to close the gap between income and absorption before supply-side factors operate fully and, as a result, some growth setbacks should normally be expected during the program. This said, the short-run growth impact of IMF-supported programs is assessed against a counterfactual of no Fund engagement. As such, the estimation aims at capturing the causal effect on growth of IMF-supported programs, using an approach designed to address possible sample selection bias.

[30] Variance ratios would be even smaller if calculated against total sample variation before country and vintage fixed effects.

95. Assessing how the growth performance of countries during IMF-supported programs has fared relative to a counterfactual of no Fund engagement is empirically challenging fundamentally because the relevant counterfactuals—i.e., growth outcomes in the absence of the program—are unknowable. Further, assessment is vulnerable to sample selection bias because IMF-supported programs are arranged only for countries with actual or prospective balance of payments problems which require policy adjustment whether or not the IMF is involved and because IMF-supported programs are an outcome of negotiation between a country and the IMF and, hence, a deliberate choice of both parties.

96. A select review of the literature indicates that evidence on the short-run growth impact of IMF-supported programs is mixed at best (Appendix V). Recent studies using a propensity scoring approach to address sample selection bias find a positive impact on growth of IMF-supported programs—see, for example, Bas and Stone (2014), Bal Gunduz (2016), and Bird and Rowlands (2017). In contrast, earlier studies point to either a negative or ambiguous effect of IMF-supported programs on growth.[31] Many factors may explain inconclusive findings of the literature, including different samples and empirical strategies adopted across studies.

Overall Impact of Engagement in an IMF-Supported Program

97. The analysis in this section attempts to correct for sample selection bias by using a recently developed statistical approach which is explained in greater detail in Appendix IV. The approach used is based on a further elaboration of the propensity scoring approach and involves modeling participation in an IMF-supported program ("treatment" model) and growth outcomes of the program ("outcome" model). It is assessed to be doubly robust in the sense that the estimator is unbiased unless both treatment and outcome models are mis-specified.

98. In this approach, the short-run growth impact of IMF-supported programs relative to a counterfactual of no Fund engagement is identified as the average treatment effect (ATE). The estimation is conducted not only for all countries in the sample but also separately for PRGT-eligible countries (PRGT countries in short) and PRGT-noneligible countries (GRA countries in short) in consideration of large differences in income level, economic structure and capacity between GRA and PRGT countries. More specifically, the growth impact of PRGT programs is estimated for the sample of PRGT countries while the growth impact of GRA programs is estimated for the sample of GRA countries.

99. Table 14 presents the results for the ATE on growth of IMF-supported programs estimated for 152 countries in total (92 GRA and 60 PRGT countries) over the period of 2008–19.[32] The estimated ATE is positive and highly significant in all cases. For completed programs (columns (1)-(3)), the engagement in an IMF-supported program is estimated to raise annual growth, relative to a counterfactual of non-participation, on average by about 0.7 percentage

[31] See Appendix V for the studies which find a negative growth impact.

[32] Appendix Table AIV.1 presents fuller results.

point in all countries, and 0.7 percentage points and 1.1 percentage points in GRA and PRGT countries, respectively.[33] Estimated growth gain is larger for PRGT programs than for GRA programs. If the program sample is expanded to include both completed and off-track programs (columns (4)-(6)), estimated growth gains are smaller in both GRA and PRGT countries, implying that program completion matters for short-run growth benefits of IMF-supported programs.

Table 14. Short-Run Growth Impact of IMF-Supported Programs: 2008–19

	Panel A. Completed Programs			Panel B. Completed and Off-track Programs		
	(1) ALL	(2) GRA	(3) PRGT	(4) ALL	(5) GRA	(6) PRGT
ATE of IMF-supported programs	0.724***	0.662***	1.050***	0.572***	0.314***	0.702***
No. of observations	1523	929	577	1565	954	594
No. of program observations	254	67	170	296	92	187
No. of non-program observations	1269	862	407	1269	862	407
R^2 of outcome model	0.572	0.661	0.441	0.576	0.663	0.449
1st stage model: CBS	Yes	Yes	Yes	Yes	Yes	Yes

Source: IEO staff estimates.
Note: Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

100. Overall, these results provide strong support for short-run growth benefits of IMF-supported programs relative to the counterfactual of not engaging in an IMF-supported program and underscore the importance of program completion in consolidating growth benefits.

Impact of Structural Conditionality in IMF-Supported Programs

101. Structural conditionality is an important aspect of IMF-supported programs which helps distinguish program periods from non-program periods. While structural reforms would normally take time to affect growth as discussed by IMF (2019c), structural conditionality may also affect short-run growth in the program context particularly if they help to boost investor or donor confidence by signaling program countries' commitment to reforms to improve economic efficiency and social inclusion. To assess whether structural conditionality has mattered for short-run growth gains, the analysis examines how structural conditions (SCs) affect the ATE on short-run growth of IMF-supported programs.

[33] Three recent studies report similar positive impact on growth. While these studies also control for selection bias, their sample periods and program types have no or limited overlap with our sample, which may explain the differences in estimates. Using a mixed sample of GRA and PRGT programs for 104 countries over 1970–2008, Bas and Stone (2014) find that the average growth impact is on the order of 1.4–3.5 percentage points and rises steadily for long-term users. Bal Gunduz (2016) reports an average growth impact of 0.4 percentage points for PRGT programs addressing immediate BOP needs in 55 LICs over 1980–2010. She finds that the impact on change in growth becomes significant and rises to 1.5–3.5 percentage points for LICs facing substantial macroeconomic imbalances or large exogenous shocks. For programs with 66 LICs over 1989–2008, Bird and Rowlands (2017) report a significant growth impact of 1.0–1.7 percentage points for concessional programs up to two years after approval but significantly negative effects for non-concessional programs. They also find growth impact tends to be stronger the weaker are initial conditions. See also Table AVI.1 in Appendix VI.

102. The analysis uses the average score index of SCs developed by Kim and Lee (2021). Specifically, each structural condition is scored between 0 and 1 for implementation status, depth and growth orientation, respectively, and composite scores are constructed by a product of two or more individual scores. The average score index for each program is then constructed by averaging individual or composite scores across SCs in the program. ASCI and ASCID stand for the average score of implementation status and the average composite score of implementation status and depth, respectively. The regression sample is then split into two subsamples depending on whether ASCI or ASCID above ("high") or below ("low") the corresponding cross-country mean.

103. Table 15 reports the estimation results for four subsamples divided based on the two average SC scores. The results in all panels provide good support for the role played by SCs and the importance of their quality in determining short-run growth gains of IMF-supported programs. Specifically, the difference in growth gains between strong and weak compliance is smaller at about 0.06 percentage points for PRGT programs but significantly larger at 1.2 percentage points for GRA programs. The results for the average composite scores paint the broadly same picture about the role of SCs and the importance of the quality of SCs in producing growth gains—namely, the difference in growth gains between low and high depth SCs is larger for GRA programs than for PRGT programs.

104. It is notable that low ASCI and ASCID are both associated with short-run growth loss, and not growth gain, in GRA programs (columns (1) and (3), panel B). This finding, together with larger growth gains from high depth SCs, provides some support for the confidence effect of SCs which would be particularly important in GRA program countries with market access. Strong implementation of high depth SCs would signal firm commitments of program countries for durable recovery and help boost investor confidence, ease external financing constraints and ultimately boost growth.[34]

105. In sum, the analysis suggests that IMF-supported programs have in general helped to boost growth in program countries, relative to a counterfactual of no Fund engagement. The estimated growth gains are on average larger for PRGT programs compared to GRA programs. The positive impact on short-run growth of IMF-supported programs is amplified when associated with strong implementation of high depth SCs, providing support for the confidence effect of SCs particularly in GRA programs.

[34] In PRGT program countries, Fund engagement itself is viewed as the most critical factor that helps boost donor confidence. This may explain generally smaller differences in growth gains between low and high SC scores (ASCI and ASCID) in PRGT programs than in GRA programs in Table 15.

Table 15. Short-Run Growth Impact of IMF-Supported Programs by Quality of SCs

Panel A. ALL Programs				
	(1) Low ASCI	(2) High ASCI	(3) Low ASCID	(4) High ASCID
ATE of IMF-supported programs	0.231***	0.761***	0.352***	0.782***
No. of observations	1387	1447	1416	1418
No. of program observations	118	178	147	149
No. of non-program observations	1269	1269	1269	1269
R^2 of outcome model	0.574	0.599	0.579	0.601
Average ASCI of treated units	0.669	0.867	0.359	0.503
1st stage model: CBPS	Yes	Yes	Yes	Yes
Panel B. GRA Programs				
	(1) Low ASCI	(2) High ASCI	(3) Low ASCID	(4) High ASCID
ATE of IMF-supported programs	-0.370***	0.874***	-0.016***	0.923***
No. of observations	905	911	904	912
No. of program observations	43	49	42	50
No. of non-program observations	862	862	862	862
R^2 of outcome model	0.659	0.667	0.662	0.662
Average ASCI of treated units	0.756	0.906	0.389	0.516
1st stage model: CBPS	Yes	Yes	Yes	Yes
Panel C. PRGT Programs				
	(1) Low ASCI	(2) High ASCI	(3) Low ASCID	(4) High ASCID
ATE of IMF-supported programs	0.472***	0.530***	0.646***	0.784***
No. of observations	487	514	503	498
No. of program observations	80	107	96	91
No. of non-program observations	407	407	407	407
R^2 of outcome model	0.435	0.551	0.445	0.558
Average ASCI of treated units	0.641	0.845	0.345	0.485
1st stage model: CBPS	Yes	Yes	Yes	Yes

Source: IEO staff estimates.
Notes: The results are based on the expanded sample that includes both completed and off-track programs. Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

XI. POST-PROGRAM GROWTH IMPACT OF IMF-SUPPORTED PROGRAMS

106. This section assesses the post-program growth impact of IMF-supported programs. Of particular interest is to assess how macroeconomic stabilizations and reforms implemented during the program have affected post-program *potential* growth. Our focus on potential growth is motivated by two considerations. First, in conceptual terms, the potential growth rate would be considered more appropriate to capture the slow-moving medium-term effects on growth of stabilizations and reforms implemented during the program. Second, use of the potential growth rate, which is corrected for cyclical variation associated with macroeconomic policies and shocks, should help to produce sharper estimates of the medium-run growth benefits of stability gains and reforms achieved in the program context.

107. In the regression analysis, the dependent variable is the post-program average of PTGBD which is discussed in Section VII. Specifically, the dependent variable is denoted by PTGBD(m) where m refers to the length of the post-program period considered. By this metric, the post-program growth performance is measured free of cyclical variation in growth arising from short-run domestic policies during the post-program period and the influence of exogenous external factors as well. The regression analysis discussed below considers 3 years (m=3) for the duration of the post-program period, mainly because of limited data availability.[35]

108. A range of explanatory variables are considered to capture stabilization outcomes of the program, including the cumulative changes during the program in the debt/GDP ratio (ΔEDY for external debt and ΔEDY for public debt), inflation ($\Delta\pi$), government revenue (ΔGREVY), public investment (ΔPUBINVY), social spending (ΔSOCIALY) and the real effective exchange rate (ΔREER).[36] In addition, debt operations undertaken in the context of the program are considered as part of stabilization outcomes and represented in the regressions by a dummy variable which takes 1 if the program has involved market debt operations and 0 otherwise.[37] Admittedly, use of a simple dummy variable is too coarse to reflect the complexity and diverse modality of debt operations across countries, but there have been too few such operations to use a more differentiated approach. As in the previous section, SC scores are used to capture reform implementation during the program. For other control variables, the regression analysis considers the government effectiveness index (GEI) and various dummies for program and country characteristics.

109. Two caveats are in order before discussing the estimation results. First, the estimation sample even for *m*=3 is small enough to cover 54 programs at most out of 132 programs included in the evaluation sample. Thus, the estimation results may not be fully representative of the average results of the programs in the evaluation sample. Second, the 3-year horizon is arguably too short to make a robust assessment of the medium-run growth impact of

[35] The post-program period of three years or longer is not well defined for many programs in the sample for several reasons: (1) some programs were followed by successor programs in less than a year or two, for which no post-program period of three years or longer can be defined; and (2) some other programs were completed only recently so that no post-program data are available for three years or longer. The regression sample is further limited by the fact that some programs quickly went off track in less than a year for which no annual frequency data are available for stabilization outcomes or reform implementation during the program.

[36] Use of cumulative change is motivated by the need to estimate the impact on post-program growth of the program. Since more stabilization could be achieved in programs of longer duration, use of cumulative change may be unfair to programs of shorter duration. In the end, however, what would matter for post-program growth would be total achievement of the program and not average achievement per program year. The same logic is applied to the measurement of structural reforms.

[37] In the evaluation sample, there are 11 programs in total that involved market debt operations in the program context, some of which are not included in the regression sample due to missing data for other explanatory variables.

stabilization gains and, particularly, structural reforms implemented during the program.[38] Given these caveats, the regression results should be considered suggestive rather than conclusive.

110. Tables AVI.4 and AVI.5 in Appendix VI present the regression results where SC score indices are considered collectively. Specifically, SCI is the aggregate implementation score while ASCD and ASCG are average depth and growth-orientation scores, respectively, as defined in the previous section.[39] The specification with interaction terms—i.e., ASCD*SCI and ASCG*SCI—is motivated to examine how the quality of SCs, and not just their volume and implementation (which are captured by SCI), would matter for growth benefits.

111. Overall, the regression results provide good support for medium-run growth benefits of both stabilization and reform efforts during the programs. As to the post-program growth impact of stabilization outcomes, debt reduction—both public and external—during the program has affected post-program potential growth positively and statistically significantly. Growth-friendly fiscal adjustment during the program also appears to have produced lasting growth benefits as increased public investment and social spending during the program are found to have positive and significant impact on post-program growth. While the growth impact of revenue mobilization is found to be slightly negative, but the estimated coefficient is not statistically significant. This result may reflect that positive benefits of higher revenue mobilization in terms of reducing the deficit and containing debt are accounted for elsewhere in the regression.

112. Market debt operations undertaken in the program context are found to have a negative and statistically significant impact on post-program growth, particularly in the specification where public debt reduction is used to capture debt outcomes of the program (see Table AVI.4). The results for debt operations may need to be interpreted with caution given that the dummy variable is too coarse to adequately capture the diverse modality and coverage of debt operations across countries. This said, a plausible interpretation for the negative coefficient of debt operations is that it may in fact capture more of the lingering effect on growth of macro-financial conditions and investor attitudes specific to programs with debt operations—such as increased borrowing cost and reduced market access after debt operations and lost investor confidence—rather than the effect of debt operations itself because the amount of debt reduction during the program is already controlled for in the regression.

113. The post-program growth impact of SCs is also positive and significant. Specifically, the coefficient of SCI is negative and significant but small in magnitude while the coefficients of the two interaction terms, particularly the interaction term of SCI and the average depth score (ASCD*SCI), are always positive, larger in magnitude than that of SCI, and highly statistically

[38] IMF (2019c) finds that structural reforms affect growth with significant time lag on the order of 5–7 years.

[39] In the regression analysis, ASCD and ASCG are measured relative to their respective possible minimum score of 0.33 given that the scale used to score depth and growth orientation is 0.33 for low, 0.66 for medium and 1.00 for high. See Kim and Lee (2021) for further technical details.

significant in many cases. This result confirms that the quality of SCs, especially the depth, matters significantly for growth benefits.[40] The negative coefficient of SCI captures the growth impact of SCs with lowest depth and growth orientation given that the two interaction terms of SCI captures the quality of implemented SCs, relative to the lowest possible depth and growth orientation. As such, the negative coefficient of SCI suggests that implementing too many low-quality SCs could even harm growth.

114. Taken together, these findings provide some support for medium-run growth benefits of stabilization and structural reforms implemented in the program context. They also suggest that the depth and growth orientation of SCs, not just volume and implementation, matter for growth benefits.

XII. CONCLUSIONS

115. Key findings from the cross-country analyses in this paper are summarized as follows.

116. ***Sustainability and growth considerations in initial program design***. Sustainability and growth considerations were well incorporated in initial program design, especially in programmed fiscal policy, in GRA programs but less clearly in PRGT programs. In GRA programs, programmed fiscal primary balance reacts positively to lagged debt ratio (for sustainability) and the output gap (out of growth considerations), suggesting that programmed fiscal policy is counter-cyclical in nature. In contrast, no statistically significant fiscal reaction is detected in PRGT programs. In addition, fiscal reaction to lagged debt ratio appears to be nonlinear, especially in GRA programs, with initial positive reaction being tapered off after debt ratio reached about 130 percent of GDP or higher.

117. ***Sustainability and growth considerations in program adaptation.*** Another form of fiscal reaction is examined to assess how sustainability and growth considerations were incorporated in program adaptation. The analysis finds that in both GRA and PRGT programs, fiscal projections were updated balancing sustainability and growth considerations. Specifically, fiscal projections (i.e., projections for fiscal adjustment) tended to revise downwards in response to growth shortfalls and upwards in response to adjustment shortfalls. Sustainability considerations were generally stronger in GRA programs than in PRGT programs.

118. ***Realism of macroeconomic framework in program design***. The macroeconomic framework in program design is found to incorporate a short-run tradeoff between fiscal adjustment and growth, but no such tradeoff has been modeled for CA adjustment consistent

[40] Although not reported, alternative specifications are considered as sensitivity check where three composite SC score indices—SCI, SCID (implementation and depth), and SCIDG (implementation, depth and growth-orientation)—enter the regression equation individually rather than collectively. The results are broadly supportive of the importance of the quality of SCs in producing post-program growth gains. Specifically, SCIDG is found to have the largest post-program growth impact while SCI has the least impact. Moreover, the growth impact of SCI is never statistically significant while those of SCID and SCIDG are in many cases.

with theory. Initial program assumptions on fiscal multipliers were on average realistic in GRA programs in the sense of being close to multipliers estimated from actual data, but updated program assumptions were less realistic than initial assumptions. In PRGT programs, the opposite seems to be the case—initial program assumptions understated actual multipliers while updated program assumptions were closer to actual multipliers than initial assumptions. In both GRA and PRGT programs, program assumptions on revenue multipliers seem to have been updated toward greater realism while the opposite is true for expenditure multipliers particularly in GRA programs.

119. ***Source of growth optimism***. The analysis finds that modeling errors on the relationship between fiscal adjustment and growth in GRA programs, especially in other GRA programs than crisis programs, have been an important source of optimism bias in program growth projections. Variance decomposition results indicate that in GRA programs other than crisis programs, fiscal projections alone explain about 30 percent of total sample variation in growth forecast errors after country and vintage fixed effects. At the aggregate level, the contribution of modeling error to growth optimism is significantly larger in program periods than in non-program periods and, within program periods, far larger in GRA programs, particularly in GRA programs other than crisis programs, than in PRGT programs. The contribution of fiscal projections dominates that of CA projections, which is consistent with the findings about program assumptions on fiscal multipliers in Section VI.

120. ***Growth outcomes against benchmark***. IMF-supported programs have fared relatively well in terms of growth if assessed relative to a common benchmark that corrects for the influence of external factors and the difference in historical trend growth. Negative and statistically significant deviations from the benchmark are found only in 12 percent of the program sample. GRA programs were populated dominantly with negative deviations, of which one-third are statistically significant while PRGT programs dominantly with positive deviations with about 20 percent being statistically significant.

121. ***Short-run growth impact of IMF-supported programs***. The analysis finds that IMF-supported programs have helped boost growth in program countries, relative to a counterfactual of no Fund engagement, based on a new technique that seeks to correct for selection bias. The estimated average growth-boosting effect of IMF-supported programs is positive, highly significant and on the order of 0.7 percentage points and 1.1 percentage points for completed GRA and PRGT programs, respectively. The analysis also finds that growth gains from IMF-supported programs are generally larger the higher is the depth of SCs, and that growth gains were more sensitive to the quality of SCs in GRA programs than in PRGT programs. These findings seem to point toward the confidence or signaling effects of SCs particularly in GRA programs where strong implementation of high depth SCs could signal firm commitments of program countries to durable recovery, which in turn boosts investor confidence and helps to ease external financing constraints.

122. ***Post-program growth impact of IMF-supported programs***. The analysis finds that both stabilization and reform implementation during the program have affected post-program potential growth positively and statistically significantly. Debt reduction (both public and external) during the program is found to have positive and significant impact on post-program growth. Growth-friendly fiscal adjustment is also found to have lasting growth benefits beyond the program horizon, particularly increased public investment and social spending. The regression analysis indicates a negative (and statistically significant in some specifications) impact of debt operations on post-program growth. Given that the amount of debt reduction is already controlled for in the regression analysis, the negative impact of debt operations seems to reflect more of the lingering effect on growth of macro-financial conditions and investor attitudes specific to programs with debt operations—such as increased borrowing cost and reduced market access after debt operations and lost investor confidence particularly in case of sovereign default—rather than the effect of debt operations itself. Structural conditions (SCs) implemented during the program are also found to have a positive impact on post-program growth with the growth impact being stronger the greater is their depth and growth-orientation, suggesting that the quality of SCs matter for post-program growth benefits of IMF-supported programs.

APPENDIX I. IMF LENDING ARRANGEMENTS: SEPTEMBER 2008–MARCH 2020

Dept.	Country	GRA	Approved	Dept.	Country	PRGT	Approved
AFR	Angola[1]	SBA	2009	AFR	Benin[1]	ECF	2010
AFR	Seychelles[2]	SBA	2008	AFR	Burkina Faso[1]	ECF	2010
AFR	Seychelles[1]	EFF	2009	AFR	Burkina Faso[1]	ECF	2013
AFR	Seychelles	EFF	2014	AFR	Burundi[1]	ECF	2012
APD	Mongolia	SBA	2009	AFR	Cape Verde	PSI	2010
APD	Sri Lanka[1]	SBA	2009	AFR	Central African Republic[2]	ECF	2012
EUR	Albania	EFF	2014	AFR	Central African Republic	ECF	2016
EUR	Belarus[2]	SBA	2009	AFR	Chad[1, 2]	ECF	2014
EUR	Bosnia and Herzegovina[1]	SBA	2009	AFR	Comoros[1]	PRGF	2009
EUR	Bosnia and Herzegovina[1]	SBA	2012	AFR	Congo, Democratic Republic of	PRGF	2009
EUR	Cyprus	EFF	2013	AFR	Congo, Republic Of	PRGF	2008
EUR	Greece[2]	SBA	2010	AFR	Cote d'Ivoire[2]	PRGF	2009
EUR	Greece[2]	EFF	2012	AFR	Cote d'Ivoire[1]	ECF	2011
EUR	Hungary[1]	SBA	2008	AFR	Ethiopia[1]	ESF	2009
EUR	Iceland[1]	SBA	2008	AFR	Gambia[2]	ECF	2012
EUR	Ireland	EFF	2010	AFR	Ghana	PRGF	2009
EUR	Kosovo, Republic of	SBA	2010	AFR	Ghana[1]	ECF	2015
EUR	Kosovo, Republic of	SBA	2012	AFR	Guinea[1]	ECF	2012
EUR	Kosovo, Republic of[1]	SBA	2015	AFR	Guinea-Bissau	ECF	2010
EUR	Latvia[1]	SBA	2008	AFR	Guinea-Bissau[1]	ECF	2015
EUR	Portugal[1]	EFF	2011	AFR	Kenya	ECF	2011
EUR	Romania[2]	SBA	2009	AFR	Kenya*[1, 2]	SBA-SCF	2015
EUR	Romania[1]	SBA	2011	AFR	Kenya*[1]	SBA-SCF	2016
EUR	Romania	SBA	2013	AFR	Lesotho[1]	ECF	2010
EUR	Serbia, Republic of[1]	SBA	2009	AFR	Liberia[1]	ECF	2012
EUR	Serbia, Republic of	SBA	2011	AFR	Madagascar	ECF	2016
EUR	Serbia, Republic of	SBA	2015	AFR	Malawi	ESF	2008
EUR	Ukraine[2]	SBA	2008	AFR	Malawi[2]	ECF	2010
EUR	Ukraine	SBA	2010	AFR	Malawi[1]	ECF	2012
EUR	Ukraine[2]	SBA	2014	AFR	Mali[2]	ECF	2011
EUR	Ukraine[2]	EFF	2015	AFR	Mali[1]	ECF	2013
MCD	Armenia[2]	SBA	2009	AFR	Mauritania[1]	ECF	2010
MCD	Armenia[1]	EFF	2014	AFR	Mozambique[1, 2]	PSI	2010
MCD	Egypt	EFF	2016	AFR	Mozambique	PSI	2013
MCD	Georgia[1]	SBA	2008	AFR	Mozambique	SCF	2015
MCD	Georgia[2]	SBA	2014	AFR	Niger[1]	ECF	2012
MCD	Iraq[1]	SBA	2010	AFR	Rwanda[1]	PSI	2010
MCD	Iraq	SBA	2016	AFR	Rwanda[1]	PSI	2013
MCD	Jordan	SBA	2012	AFR	Rwanda[1]	SCF	2016
MCD	Jordan[1]	EFF	2016	AFR	Sao Tome & Principe	PRGF	2009
MCD	Pakistan[1]	SBA	2008	AFR	Sao Tome & Principe	ECF	2012
MCD	Pakistan[1]	EFF	2013	AFR	Sao Tome & Principe[1]	ECF	2015

Dept.	Country	GRA	Approved	Dept.	Country	PRGT	Approved
MCD	Tunisia[1]	SBA	2013	AFR	Senegal[1]	PSI-ESF	2008
WHD	Antigua and Barbuda	SBA	2010	AFR	Senegal[1]	PSI	2010
WHD	Costa Rica	SBA	2009	AFR	Senegal[1]	PSI	2015
WHD	Dominican Republic	SBA	2009	AFR	Sierra Leone	ECF	2010
WHD	El Salvador[2]	SBA	2009	AFR	Sierra Leone[1]	ECF	2013
WHD	El Salvador	SBA	2010	AFR	Tanzania	PSI	2010
WHD	Guatemala	SBA	2009	AFR	Tanzania[1]	SCF	2012
WHD	Jamaica	SBA	2010	AFR	Tanzania[1]	PSI	2014
WHD	Jamaica[2]	EFF	2013	AFR	Uganda[1]	PSI	2010
WHD	Jamaica	SBA	2016	AFR	Uganda[1]	PSI	2013
WHD	St. Kitts and Nevis	SBA	2011	APD	Bangladesh[1]	ECF	2012
WHD	Suriname[2]	SBA	2016	APD	Maldives*	SBA-ESF	2009
				APD	Solomon Islands	SCF	2010
				APD	Solomon Islands	SCF	2011
				APD	Solomon Islands[1]	ECF	2012
				EUR	Moldova*[1]	EFF-ECF	2010
				EUR	Moldova*	EFF-ECF	2016
				EUR	Tajikistan[1]	PRGF	2009
				MCD	Afghanistan, I. S. of	ECF	2011
				MCD	Afghanistan, I. S. of[1]	ECF	2016
				MCD	Armenia[2]	PRGF	2008
				MCD	Armenia*[1]	EFF-ECF	2010
				MCD	Djibouti[1]	PRGF	2008
				MCD	Georgia*	SBA-SCF	2012
				MCD	Kyrgyz Republic	ESF	2008
				MCD	Kyrgyz Republic[1]	ECF	2011
				MCD	Kyrgyz Republic	ECF	2015
				MCD	Yemen, Republic of[2]	ECF	2010
				MCD	Yemen, Republic of[2]	ECF	2014
				WHD	Grenada	ECF	2010
				WHD	Grenada	ECF	2014
				WHD	Haiti[1]	ECF	2010
				WHD	Haiti[2]	ECF	2015
				WHD	Honduras*	SBA-SCF	2010
				WHD	Honduras*	SBA-SCF	2014

* Indicates GRA-PRGT blended arrangements.

[1] Completion delayed or program extended but not completed (Pakistan 2008).

[2] Cancelled.

APPENDIX II. ADJUSTMENT AND GROWTH IN PROGRAM DESIGN: AN ANALYTICAL FRAMEWORK

This appendix illustrates a simple analytical framework which explains how adjustment and growth would be determined in program design of IMF-supported programs and has guided the assessment of various aspects of the program design and growth outcomes in this report.

The determination of the balance between fiscal adjustment and growth in program design can be succinctly illustrated by using the public debt dynamics equation, which is given by

(1) $$\Delta d_1 = (r_0 - g_1)d_0 - s_1$$

where d denotes the public debt-to-GDP ratio, r the real effective interest rate on debt, g the real GDP growth rate, and s the fiscal primary balance as a share of GDP. Subscripts 0 and 1 refer to the year of program initiation and the year of program completion, respectively.

The magnitude of targeted debt reduction, $-\Delta d_1 \geq 0$, is assumed to be determined by considerations on debt sustainability, available financing and the country's adjustment capacity, among others. Equation (1) can be rewritten into the following relationship between growth and fiscal adjustment, which is labeled by DS:

(2) DS: $$g_1 = q - \Delta s_1/d_0, \qquad q' > 0$$

where $q = r_0 - \Delta d_1/d_0 - (s_0/d_0)$, and $\Delta s_1 = s_1 - s_0$ denotes fiscal adjustment. DS schedule represents all pairs of $\{g_1, \Delta s_1\}$ consistent with a given amount of programmed debt reduction (Δd_1) and shifts outward as the programmed debt reduction increases.

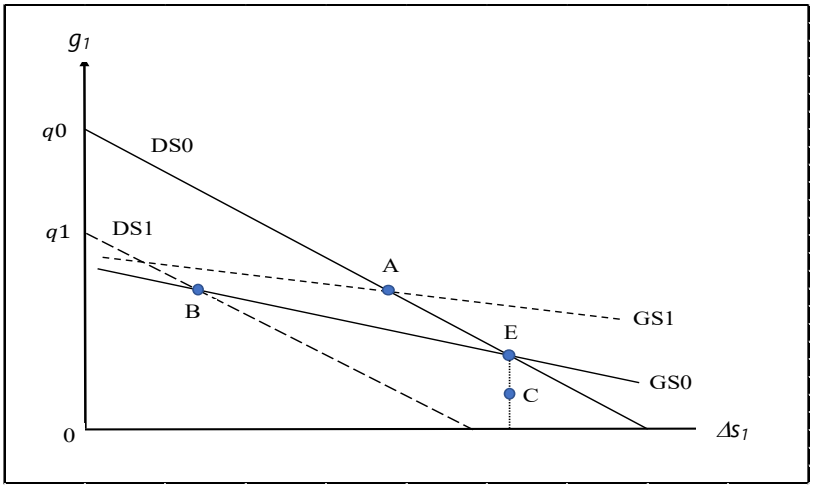
Not all points on DS schedule are attainable, however, because DS schedule does not represent the equilibrium relationship between fiscal adjustment and growth. Abstracting from uncertainty, the equilibrium behavioral relationship between g_1 and Δs_1, which is labeled by GS, is characterized by

(3) GS: $$g_1 = g^p + g^c(\Delta s_1, z), \qquad g_1^c < 0, \quad g_2^c > 0$$

where g^p and g^c capture the potential and cyclical components of the growth rate, respectively. It is assumed that g^c depends negatively on fiscal adjustment (Δs_1) so that GS schedule is downward sloping, and positively on growth-supporting measures (z), while g^p is predetermined in the short run.

Program projections of $\{g_1, \Delta s_1\}$ should be consistent with both DS and GS schedules, implying that they are determined at the intersection of DS and GS schedules as illustrated in Figure AII.1. In the Figure, without loss of generality, DS0 is drawn for $\Delta d_1 = 0$ while GS0 is drawn for $z = 0$ (i.e., no growth-supporting measures). In equilibrium, point E determines the balance between adjustment and growth envisaged in program design.

Figure AII.1. Determination of Growth and Fiscal Adjustment in Program Design: An Illustration



Role of Growth-Supporting Measures

IMF-supported programs include a range of growth-supporting measures to enhance growth while delivering needed adjustment. Prime examples are (1) protection of public investment and social spending financed by tax increases, (2) exchange rate flexibility, (3) structural reforms, and (4) debt operations. The first three measures operate primarily by affecting GS schedule, while debt operations operate primarily by affecting DS schedule.

- *Protection of public investment/social spending and exchange rate flexibility* may help reduce the short-run fiscal multiplier, implying that GS schedule becomes flatter than otherwise and/or shifts upward (as shown by GS1 for $z > 0$ in Figure AII.1). Consequently, with these growth-supporting measures, program projections of $\{g_1, \Delta s_1\}$ are determined at point A with higher growth and less adjustment than point E.

- *Structural reforms* are intended primarily to enhance medium to long run growth by improving the potential growth rate (g^p). But they can also affect short-run growth if they help boost investor/donor confidence, which underpins the catalytic role of IMF-supported programs. GS schedule would shift upward (along with an increase in g^p) over the post-program period and/or during the program period through the confidence effect. Thus, point A would be a good characterization of the impact of structural reforms on post-program or short-run growth.

- *Debt operations* can help reduce the adjustment need over the program period and enhance growth by reducing the level of debt (debt restructuring) or debt services falling due (debt reprofiling). Debt operations can be modeled as an inward shift of DS schedule (from DS0 to DS1 in Figure AII.1), in which case program projections of $\{g_1, \Delta s_1\}$ are determined at point B with higher growth and less adjustment than point E.

The stylized framework discussed above would need to be qualified because some of the assumptions may not hold in practice depending on country circumstances. In this respect, several caveats are in order. First, exchange rate flexibility may not be warranted for some program countries where public debt involves foreign currency debt, due to concerns on the adverse balance sheet effect (which would correspond to a shift in DS schedule) and/or unstable inflation expectations. Second, GS schedule may not intersect DS schedule within an economically and politically feasible set of $\{g_1, \Delta s_1\}$, in which case debt operations may be called for (to shift DS schedule inward until feasibility constraints are met). Third, debt operations could be contractionary during the negotiation stage although expansionary once completed. Fourth, some structural reforms could be contractionary in the short run or over the implementation phase if political resistance to reforms leads to heightened economic uncertainty.

Role of Growth Uncertainty and Modeling Errors

Program outcomes do deviate from program projections due to unforeseen shocks including slippages in program implementation as well as modeling errors in program design (e.g., faulty assumptions on the fiscal multiplier). Unforeseen shocks to growth shifts GS schedule upward or downward and can be modeled as follow:

(4) $$g_1 = g^p + g^c(\Delta s_1, z) + \varepsilon_1, \quad E(\varepsilon_1) = 0, \ \ \varepsilon_1 \in [-\overline{\varepsilon}, \ \overline{\varepsilon}]$$

In case of a negative shock ($\varepsilon_1 < 0$), for instance, projected growth at point E will turn out to be optimistic while debt sustainability may deteriorate with an increase in the debt ratio (point C in Figure AII.1). Since the shock is zero on average, however, any systematic bias in program projections for growth, if occurred, should be attributed primarily to modeling errors in program design. The effect of modeling errors could be analyzed by comparing point A and E in the Figure if program design is based on GS1 when GS0 represents the true behavioral relationship.

APPENDIX III. BOP NEED DECOMPOSITION: METHODOLOGY AND RESULTS

Following the methodology used by the 2018 ROC, the BOP need decomposition is undertaken for the case study sample of 40 programs arranged for 17 countries over the period of 2008–19, where data used for the decomposition are program projections in initial program design.

Methodology

The BOP need decomposition involves an estimation of the counterfactual BOP need—i.e., BOP need assuming continuation in the pre-program balance of payments—and how the estimated BOP need is to be met by CA adjustment and different sources of financing.

Counterfactual current account deficits (CADs) are calculated on the before-grant basis by assuming that the CAD/GDP ratio at T-1 is maintained throughout the program period—i.e., applying the CAD/GDP ratio at T-1 to projected GDP over the program period. A similar process is followed to construct counterfactual financing by assuming that the KFB/GDP ratio at T-1 (or at T if smaller than T-1) is maintained throughout the program period where KFB denotes the sum of the capital account (before official transfers) and financial account balances. The counterfactual BOP need is then calculated as the difference between counterfactual CADs and counterfactual financing.

In the BOP need decomposition, CA adjustment is measured as the difference between the cumulative total of counterfactual and programmed CADs over the program period as a share of GDP at time T. Data on program financing—Fund financing and exceptional financing from IFIs and bilateral official creditors—are taken from the program approval document of each program in the sample.[1] Finally, other financing is the residual component of the counterfactual BOP need after CA adjustment and program financing. It should also be noted that, for consistency with the 2018 ROC methodology, the data convention discussed in Section II is not applied in the BOP need decomposition.

The presentation of the decomposition results shown in Section III.1 deviates from the practice used by the 2018 ROC in two respects. First, the contributions of CA adjustment and financing to meeting the BOP need are shown in percent of GDP, while they are presented as a percent share of the overall BOP need in the 2018 ROC. Although intuitive, the shares in percent of total BOP need mask the significant variation in the size of overall BOP need across countries. We prefer to present the shares in percent of GDP because it makes more meaningful cross-country comparisons. Second, the results of BOP need decomposition in the 2018 ROC were presented in terms of cumulative total figures over the program period. As we are presenting the shares in percent of GDP, however, use of the cumulative total is less useful for cross-country comparison because Fund programs are of different duration across countries and program duration matters

[1] Following the methodology used by the 2018 ROC, financing from bilateral official creditors is adjusted to include grants and official transfers in the current and capital account balances, respectively.

for the cumulative total. For cross-country comparison on a more equal footing, the results of BOP need decomposition are presented on an annualized basis—i.e., average BOP need and shares of each component per program year as a share of GDP at time T.

Cross-Country Results

Table AIII.1 shows the summary results for the decomposition of the BOP need for the case study sample of 40 programs (22 GRA and 18 PRGT) arranged for 17 countries. The annualized BOP need is on average larger in GRA-supported programs (9.8 percent of GDP) than in PRGT-supported programs (8.8 percent of GDP). Within the GRA sample, exceptional access programs stand out with significantly larger total BOP need, CA adjustment and IMF financing than other GRA-supported programs.

Table AIII.1. BOP Need Decomposition: Cross-Program Comparison (In percent of GDP, Annual)

	Full Sample	GRA			PRGT
		Total	Exceptional Access Program	Crisis Program	
Total BOP Need	9.4	9.8	14.1	9.2	8.8
CA Adjustment	2.3	2.9	4.3	1.9	1.5
Financing	7.1	6.9	9.9	7.3	7.3
IMF	1.1	1.7	2.9	1.0	0.4
IFIs and Bilaterals	3.2	2.6	3.4	3.4	4.0
Other	2.7	2.5	3.6	2.9	2.8

Source: IEO staff estimates.

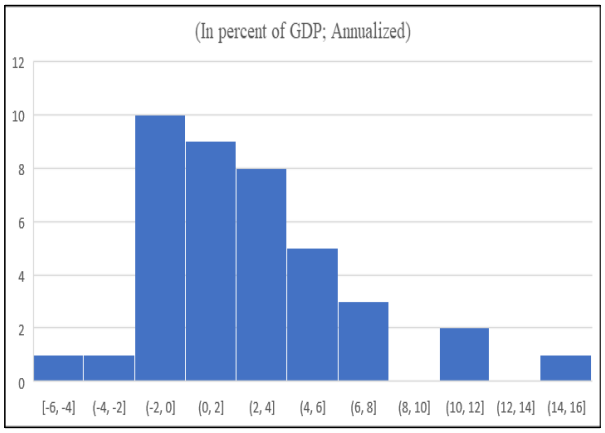
Programmed CA adjustment (relative to a counterfactual) has ranged widely from -6.0 percent to 15.3 percent of GDP per year—being negative in 12 programs in the sample, the significant majority of which are PRGT programs (Figure AIII.1, Panel A). CA adjustment per year exceeded 10 percent of GDP in three programs in the sample—Ghana (2009), Latvia (2009) and Jamaica (2010)—reflecting in large part the positive developments in exports in Ghana, significant import compression following programmed fiscal consolidation in Latvia, and the combined effects of a projected rebound in exports following real currency depreciation prior to the program and lower imports on account of weak domestic demand and fiscal retrenchment.

IMF financing has ranged between 0 percent and 4.1 percent of GDP per year (see Figure AIII.1, Panel B). In contrast to CA adjustment, the cross-country distribution of IMF financing is heavily concentrated and skewed with 23 programs out of 40 programs in the sample—or 17 programs out of 34 programs if precautionary and PSI-supported programs with no IMF financing are excluded—falling into the range between 0 percent and 1 percent of GDP.[2] Such concentration is broadly consistent with the analysis of the 2018 ROC where the Fund's access limits/norms are found to be the most important and robust determinant of program access.

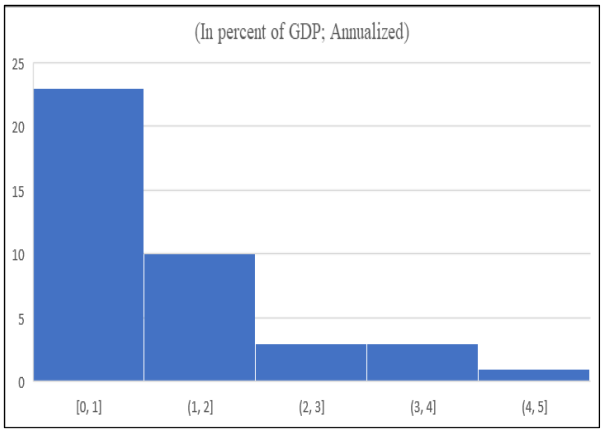
[2] Six programs in the case study sample with no IMF financing consist of four precautionary programs and two PSI-supported programs for which no access or disbursement was assumed at program approval.

Figure AIII.1. Cross-Country Distribution of CA Adjustment and IMF Financing (Case Study Sample)

A. CA Adjustment



B. IMF Financing



Source: IEO staff estimates.

APPENDIX IV. ESTIMATING SHORT-RUN GROWTH IMPACT OF IMF-SUPPORTED PROGRAMS

Estimation Method

The analysis attempts to correct for sample selection bias by using a recently developed statistical approach known as the inverse propensity weighted regression adjustment estimator (IPWRA) by Lunceford and Davidian (2004), Wooldridge (2007), and Glynn and Quinn (2010).[1] This approach has been developed to address the general problem of avoiding sample selection bias in the presence of missing data. In this paper, this approach is specifically applied to address the problem of missing counterfactual—because of the absence of the data on countries' experiences if they had chosen not to undertake an IMF-supported program.

Specifically, the approach used in our analysis involves two stages of modeling. The first stage model (or treatment model) is to estimate the probability of participating in an IMF-supported program. The second stage model (or outcome model) is to estimate the growth rate by using the data weighted by the inverse propensity obtained from the first stage model. This approach is doubly robust in the sense that the IPWRA estimator is unbiased unless both treatment and outcome models are mis-specified. It has been known to outperform the first-generation method based on propensity score which was used by Hutchison (2004), Bas and Stone (2014), Bal Gunduz (2016) and Bird and Rowlands (2017).

In our analysis on the short-run growth impact of IMF-supported programs, the first stage model is specified as:

(A.1) $$D_{it} = \gamma_i + W_{it}\Phi + \upsilon_{it}$$

where countries and time periods are indexed by subscripts i and t, respectively, D is program dummy and W is a set of relevant predictors of IMF-supported programs. This first stage model is estimated for a large sample of program and non-program countries and periods. The second stage model is given by

(A.2) $$g_{it} = \beta_i + \alpha D_{it} + X_{it}\Gamma + \varepsilon_{it}$$

where g is the growth rate, and X is a set of growth determinants. This second stage model is estimated for the combined sample of programs and non-programs while allowing a program dummy and country fixed effects in the regression equation, following the approach used by Jordà and Taylor (2016) to deal with the significantly small size of the treatment group relative to that of the control group.[2]

---

[1] The method has been used in several recent studies: Jordà and Taylor (2016) to assess the effects of fiscal consolidation on growth, and Kuvshinov and Zimmermann (2016), Asonuma and others (2019), and Atsebi and others (2019) to estimate the output and trade costs of financial crises.

[2] As a robustness check, the outcome model was estimated separately for program and non-program samples. The results are qualitatively and quantitatively similar to those reported in Table 14.

The estimated probability in the first stage, denoted by p_{it}, is used to randomize the sample used for the estimation of the second stage outcome model. Specifically, the data for the treated (i.e., program observations) and control (i.e., non-program observations) groups are weighted by $1/p_{it}$ and $1/(1-p_{it})$, respectively. Randomization is necessary to eliminate the difference in observables between program and non-program periods covered in the second stage regression and mimic a situation where IMF-supported programs are arranged randomly (see Rosenbaum and Rubin, 1983). The second stage model estimated for the program sample ($D_{it}=1$) is used to predict the growth outcomes of the whole sample. Predicted growth rates are considered as the growth rates that would have been realized if an IMF-supported program had been in place. Likewise, the second stage model estimated for the non-program sample ($D_{it}=0$) is used to predict the growth outcomes of the whole sample, which are taken as the growth rates that would have prevailed in the absence of IMF-supported programs. The average treatment effect (ATE) on growth of IMF-supported programs is then identified as the average difference between these two sets of predicted growth rates.

The first stage model is estimated by using a covariate balancing propensity score (CBPS) estimator proposed by Imai and Ratkovic (2014), which ensures a perfect covariate balancing—i.e., full elimination of the differences in characteristics between program and non-program periods. It is known to outperform the traditional logit estimators. In this respect, CBPS is better suited to identify a *causal* effect running from IMF-supported program to growth. The explanatory variables (denoted by W_{it} in equation (A.1)) included in the first stage model are as follows: fiscal primary balance/GDP (PBY), CA balance/GDP (CABY), percentage change in REER (ΔREER), percentage change in the terms of trade (ΔTOT), trading partners' growth (TP_GR), output gap (YGAP), log of public debt/GDP (ln(PDY)), log of external debt/GDP (ln(EDY)), log of domestic credit/GDP (ln(DCY)), log of per capita real GDP (ln(Y/N)), foreign reserves/imports (RESI), government effectiveness index (GEI) and a crisis dummy for debt, fiscal, banking and currency crises.

The second stage model is estimated by using OLS regressions and considers the following explanatory variables: change in fiscal primary balance/GDP (ΔPBY), change in CA balance/GDP (ΔCABY), percentage change in REER (ΔREER), percentage change in the terms of trade (ΔTOT), trading partners' growth (TP_GR), output gap (YGAP), public debt/GDP (ln(PDY)), external debt/GDP (ln(EDY)), per capita real GDP (ln(Y/N)), government effectiveness index (GEI), and a crisis dummy that captures debt, fiscal, banking and currency crises.[3] Although program financing is not included as a control variable, the inclusion of program dummy as shown in (A.2)

[3] Use of the OLS regressions may bias the coefficients of the endogenous variables such as ΔPBY, ΔCABY and ΔREER. For sensitivity check, OLS regressions are also estimated by using the lagged values of these endogenous variables. The results are broadly similar to those reported in Table 14.

helps to capture the growth impact of program financing which operates independently from policy adjustments.[4]

Results of Second Stage Regressions

The estimation of the ATE of IMF-supported program on growth is conducted for a large panel sample of 152 countries over the period of 2008–19. Table AIV.1 presents the results of second stage regressions. The results show that IMF-supported programs have on average a positive and highly significant growth effect both in GRA and PRGT countries. Also notable is that some of the control variables remain statistically significant even after the randomization of the sample with the propensity scores estimated in the first stage model, an indication of the advantages of the two-stage estimation. Specifically, output gap, percentage change in the terms of trade and trading partners' growth are positively associated with growth while crises are negatively associated with growth.

Table AIV.1. Results of Second Stage Regressions

	A. Completed Programs			B. Completed and Off-track Programs		
	(1) ALL	(2) GRA	(3) PRGT	(4) ALL	(5) GRA	(6) PRGT
ATE of IMF-supported programs	0.724***	0.662***	1.050***	0.572***	0.314***	0.702***
Change in fiscal primary balance/GDP (ΔPBY)	0.042	0.010	0.005	0.042*	-0.004	0.016
Change in CA balance/GDP((ΔCABY)	-0.035*	0.006	-0.006	-0.037*	0.010	-0.017
Percentage change in REER (ΔREER)	0.015	-0.000	-0.006	0.010	0.001	-0.006
Output gap (YGAP)	0.500***	0.364***	0.594***	0.493***	0.356***	0.625***
Debt, banking, currency or fiscal crisis	-0.806***	-1.751***	-0.378	-0.973***	-1.590***	-0.232
Public debt/GDP (ln (PDY))	0.750	0.082	1.237	0.690	0.096	0.725
External debt/GDP (ln (EDY))	-0.329	-1.836***	-0.352	-0.219	-2.002***	-0.116
Per capita real GDP (ln (Y/N))	-1.136	-1.067	-2.517**	-1.351	-1.189	-1.734
Percentage change in the terms of trade (ΔTOT)	0.034***	0.017	0.026**	0.030***	0.027*	0.031**
Trading partners' growth (TP_GR)	0.660***	0.734***	0.253**	0.672***	0.745***	0.276***
Government effectiveness index (GB)	0.468	-0.537	1.872*	0.287	-0.442	1318
No. of observations	1523	929	577	1565	954	594
No. of programs observation	254	67	170	296	92	187
No. of non-programs countries	1269	86.2	407	1269	862	407
R^2 of outcome mode	0.572	0.661	0.441	0.576	0.663	0.449
1st stage model: CBPS	Yes	Yes	Yes	Yes	Yes	Yes

Source: IEO staff estimates.
Note: Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

[4] For a given BOP gap, more program financing would be associated with on average less adjustment than otherwise. Thus, the inclusion of fiscal and CA adjustments as a control in the second stage growth regression will capture some of the growth impact of program financing that operates via reduced adjustment, provided that less adjustment leads to higher growth than otherwise. Including program dummy as well would then help to capture the growth impact of program financing beyond what is captured via adjustment variables.

CBPS vs. Logit

For sensitivity check, the first stage model is also estimated by using logit regressions, the results of which are broadly in line with those reported in Table 14. Table AIV.2 shows the covariate balancing diagnostics between program and non-program periods before and after randomization. As anticipated, the results in the table show that there are significant differences in characteristics between programs and non-period programs. For instance, countries/periods under an IMF-supported program are likely to be characterized with larger CA deficits, higher incidence of crisis and recession, higher public debt/GDP, lower domestic credit/GDP, lower reserves/imports, and lower governance effectiveness. The difference in average growth between program and non-program periods may not be exclusively attributed to the IMF-supported programs, but also to the differences in characteristics highlighted above. To capture a causal effect of IMF-supported program on growth, we use the probability of participating in a program predicted in the first stage using the CBPS estimator to fully eliminate the differences in characteristics between program and non-programs period.

Table AIV.2. Covariate Balancing Diagnostics

	(1) Observed				(II) CBPS				(III) Logit			
Variables	Mean Programs	Mean Non-Programs	Std mean diff	Var-ratio	Programs	Non-programs	Std mean diff	Var-ratio	Programs	Non-programs	Std mean diff	Var-ratio
PBY	-1.480	-0.591	-0.170	0.420	-0.990	-0.990	0.000	0.389	-1.117	-0.771	-0.067	0.402
CABY	-8.306	-2.219	-0.634	0.684	-5.767	-5.767	0.000	0.261	-6.604	-3.867	-0.287	0.429
ΔREER	-0.200	1.027	-0.189	1.098	0.555	0.555	0.000	0.565	0.613	0.688	-0.012	0.726
Crisis	0.618	0.210	0.911	1.428	0.350	0.350	0.000	1.003	0.364	0.302	0.132	1.101
Output gap	-0.226	0.637	-0.232	0.952	0.139	0.139	0.000	1.044	0.439	0.374	0.016	1.079
PDY (log.)	3.794	3.596	0.297	0.493	3.704	3.704	0.000	0.722	3.676	3.647	0.042	0.635
EDY (log.)	3.830	3.835	-0.006	0.711	3.851	3.851	0.000	1.009	3.712	3.838	-0.135	0.740
DCY (log.)	3.060	3.606	-0.487	1.583	3.320	3.320	0.000	0.662	3.161	3.456	-0.279	0.763
RGDPcap (log.)	6.252	6.568	-0.099	0.743	6.409	6.409	0.000	0.693	6.378	6.489	-0.035	0.745
RESI	0.365	0.480	-0.330	0.183	0.429	0.429	0.000	0.328	0.398	0.452	-0.164	0.233
ΔTOT	2.391	1.099	0.111	1.353	1.539	1.539	0.000	0.754	1.743	1.435	0.027	0.960
TPGR	2.976	3.146	-0.078	1.085	3.185	3.185	0.000	0.721	3.180	3.132	0.023	0.765
GEI	-0.492	-0.041	-0.570	0.585	-0.283	-0.283	0.000	0.643	-0.479	-0.162	-0.405	0.536

Source: IEO staff estimates.
Note: Rubin (2002) suggests the use of the absolute value of the standardized difference (Std-diff) as a balance measure for the first moment, where the balance is defined by absolute values below 0.25. He also suggests the use of the ratio of treated and control variances (Var-ratio) as a balance measure for the second moment, where the balance is defined by values close to 1.0, and variables are out of balance if the variance ratio is greater than 2.0 or less than 0.5.

APPENDIX V. GROWTH IMPACT OF IMF-SUPPORTED PROGRAMS: A SELECT LITERATURE REVIEW

There is a large academic literature examining the impact of IMF-supported programs on growth (and other macroeconomic variables) using various statistical analyses. A select review of the literature suggests that little consensus has thus far emerged on the growth impact of IMF-supported programs, although recent studies suggest a more positive impact on growth of IMF-supported programs than earlier studies (Table AV.1). As is well known, there are significant empirical challenges involved in identifying the growth impact of IMF-supported programs, ranging from inherent difficulties to establish relevant counterfactuals to challenges in controlling endogeneity biases in statistical analysis.

Table AV.1. Selected Studies on the Growth Impact of IMF-Supported Programs

	Impact on growth	Period	Countries	Type of programs	Methodology
No Selection Correction					
Bordo and Schwartz (2000)	-	1973–1998	14 EMs	SBA EFF	
Dicks-Mireaux, Mecagni, Schadler (2000)	+*	1986–1991	61 LICs	ESAF	
Evrensel (2002)	+*	1971–1997	91 countries	Mixed	
Lee and Rhee (2003)	+	1973-1994	88 countries	Mixed	
Selection Correction: Heckman Correction or Instrumental Variables					
Przeworski and Vreeland (2000)	-*	1970–1990	79 countries	Mixed	Heckman
Hutchison (2003)	-*	1975–1997	67 countries	SBA EFF	Heckman
Vreeland (2003)	-*	1961–1993	110 countries	Mixed	Heckman
Independent Evaluation Office (2002)	-*/-	1975–1999	82 countries	Mixed	IV
Barro and Lee (2005)	-*	1975–1999	130 countries	SBA EFF	IV
Butkiewicz and Yanikkaya (2005)	-*	1970–1999	100 countries	Mixed	IV
Easterly (2005)	0	1980–1999	20 repeat users	Mixed	IV
Dreher (2006)	-*	1970–2000	98 countries	SBA EFF	IV
Eichengreen, Gupta, and Mody (2008)	-*	1980–2003	24 countries	SBA EFF	Heckman
Marchesi and Sirtori (2011)	- *	1982–2005	128 countries	Mixed	IV
Bas and Stone (2014)	+*	1970-2008	104 countries	Mixed	Heckman
Selection Correction: Propensity Score Matching					
Hardoy (2003)	-	1970–1990	109 countries	Mixed	PSM PSM(DID)
Hutchison (2004)	+	1975–1997	25 EMs	SBA EFF	PSM; Heckman
Atoyan and Conway (2006)	+/-	1993–2002	95 countries	Mixed	PSM; Heckman
Bal Gündüz (2016)	+*	1980-2010	55 LICs	Concessional (shock)	PSM
Bird and Rowlands (2017)	+*	1989-2008	66 LICs	SBA EFF- ESAF PRGF	PSM

Sources: Steinwand and Stone (2008) and Bal Gündüz (2016).
Note: Heckman = Heckman two-step estimator for correcting selection bias; IV = Instrumental variable estimator; PSM = Propensity Score Matching; DID= Difference-in-differences; EMs = Emerging Markets; EFF = Extended Fund Facility; ESAF = Enhanced Structural Adjustment Facility; LICs = Low-Income Countries; SBA = Stand-By Arrangement. +* Significantly positive; -* Significantly negative; + Positive but insignificant; - Negative but insignificant; 0 Very close to zero.

The mixed results on the growth impact in the literature reflect differences in time periods and subsets of programs covered across studies as well as differences in empirical approaches used. Given the empirical challenges involved in establishing relevant counterfactuals, however, the differences in empirical approaches used seem to matter significantly for the mixed results in the

literature. Since 2000, correcting for selection bias has become a standard component of the analysis: most studies have applied either the Heckman two-stage methodology or instrumental variable (IV) regressions, while a handful of studies implemented propensity score matching (PSM) approach.

Each method has pros and cons. Heckman and IV regressions are best suited when unobservable factors significantly drive selection while the key challenge is to identify exclusion restrictions, i.e., finding variables strongly correlated with the participation in an IMF-supported program but not correlated with growth. PSM is best suited when selection is based on observables. However, PSM results can be tested using Rosenbaum sensitivity analysis to assess how much hidden bias can be present before results of the study begin to change. In the IV approach, as is well known, weak instruments pose a problem. Some studies used political variables as instruments, such as the size of governments, the quota at the IMF, the U.S. and European influence, and the number of veto players. Evidence is mixed in terms of their impact on participation; some find a role for U.S. influence, but limited to non-concessional lending. Others suggest that U.S. influence has an impact on other aspects of the IMF lending rather than participation, such as the size of loans, nature of structural conditionality, and record of program implementation.

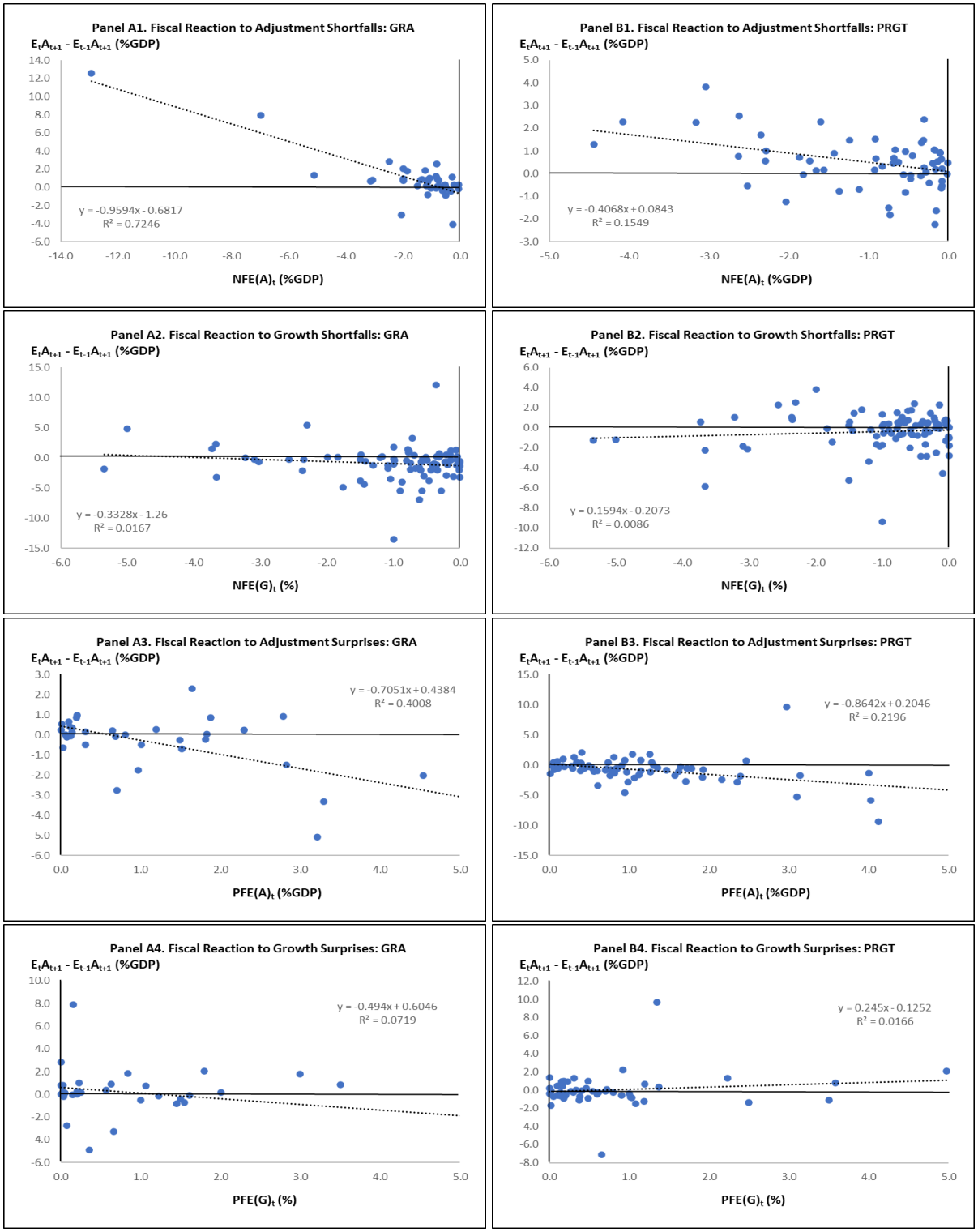
The design of the empirical work also matters. Studies using mixed samples are unlikely to estimate a strong selection model, which is critical for both Heckman and PSM approaches. In the context of participating in an IMF-supported program, focusing on a relatively homogenous subsets of programs by type (such as SBA versus ECF) and country groups (EMs versus LIDCs) significantly improves the selection model.

Three most recent studies report significantly positive impact on growth using Heckman and PSM approaches. Bas and Stone (2014) find that the average growth impact of IMF-supported programs is on the order of 1.4-3.5 percentage points and rises steadily with the cumulative number of years under programs, i.e., for longer-term IMF engagement. Bal Gunduz (2016) reports an average growth impact of 0.4 percentage points for PRGT programs and finds that the growth impact rises to 1.5-3.5 percentage points in LICs facing substantial macroeconomic imbalances or large exogenous shocks. Bird and Rowlands (2017) report a significant growth impact of 1.0–1.7 percentage points for concessional programs up to two years after approval but negative effects for non-concessional programs.

Several other studies looked at the impact of IMF-supported programs in a broader scope encompassing inclusiveness and income inequality. For example, Garuda (2000) finds that IMF-supported programs tend to improve income distribution relative to non-program countries if initial external imbalances are small or moderate while the opposite holds otherwise. Conway (2009) finds for the sample of 108 developing countries over the period of 1988–98 that income inequality depends primarily upon country characteristics, and not on participation in IMF-supported programs. In contrast, Oberdabernig (2013) shows for the sample of 86 LICs over the period of 1982–2009 that there are adverse short-run effects on poverty and income inequality of IMF-supported programs.

APPENDIX VI. SUPPLEMENTARY FIGURES AND TABLES

Figure AVI.1. Fiscal Reactions to Adjustment and Growth Forecast Errors



Source: IEO staff estimates.

Table AVI.1. Adjustment and Growth: Initial Projections and Outcomes

	Program Design					Program Outcome				
	Mean	Median	Min	Max	STD	Mean	Median	Min	Max	STD
A. GRA										
Growth	2.1	2.2	-4.9	7.7	2.5	0.7	1.7	-15.1	7.3	4.2
Fiscal Adjustment (% GDP)	1.2	0.7	-3.0	9.1	2.6	1.3	0.9	-2.6	9.0	2.8
Revenue	0.0	0.1	-2.4	1.4	0.9	0.1	0.0	-6.3	4.2	1.5
Primary Expenditure	-1.2	-0.6	-9.5	1.9	2.4	-1.2	-0.4	-10.5	2.9	3.0
CA Adjustment (% GDP)	1.5	1.3	-2.2	7.0	1.9	2.1	1.1	-4.9	13.2	3.5
Exports	0.6	0.5	-4.4	5.8	1.9	1.1	1.1	-12.0	12.6	3.9
Imports	-0.9	-0.6	-5.0	7.2	2.1	-1.0	-0.9	-11.4	8.4	4.3
Change in:										
Public Debt (% GDP)	-2.2	-0.5	-44.2	9.9	9.8	0.7	1.6	-86.0	32.7	16.4
External Debt (% GDP)	-0.9	0.1	-46.8	17.1	11.7	8.2	2.2	-36.5	121.8	26.0
B. GRA Crisis Programs										
Fiscal Adjustment (% GDP)	1.4	0.7	-2.4	9.1	2.7	1.7	0.7	-2.6	9.0	3.2
Revenue	0.0	0.2	-1.8	1.4	0.9	-0.1	0.0	-6.3	4.2	2.1
Primary Expenditure	-1.4	-1.0	-8.5	1.9	2.4	-1.8	-0.5	-10.5	1.7	3.4
CA Adjustment (% GDP)	2.0	1.6	-0.2	6.7	1.6	2.3	2.6	-3.9	11.2	2.9
Exports	0.7	0.8	-3.6	3.8	1.8	1.0	1.1	-12.0	5.9	3.8
Imports	-1.4	-0.6	-5.0	1.7	1.9	-1.4	-0.5	-10.1	5.2	4.2
Change in:										
Public Debt (% GDP)	3.5	6.3	-7.3	9.9	6.4	6.1	6.6	-13.4	26.9	7.7
External Debt (% GDP)	1.2	2.9	-36.3	17.1	12.8	12.8	5.4	-27.7	121.8	27.3
C. Other GRA Programs										
Growth	3.2	3.3	0.9	7.6	1.6	2.2	2.6	-5.6	6.4	2.8
Fiscal Adjustment (% GDP)	1.1	0.7	-3.0	8.8	2.5	1.0	0.9	-2.1	8.9	2.2
Revenue	0.0	0.1	-2.4	1.4	0.9	0.2	0.1	-1.0	1.7	0.7
Primary Expenditure	-1.1	-0.4	-9.5	1.1	2.3	-0.8	-0.4	-9.6	2.9	2.6
CA Adjustment (% GDP)	1.0	0.7	-2.2	7.0	2.0	1.8	0.6	-4.9	13.2	4.1
Exports	0.6	0.4	-4.4	5.8	1.9	1.3	1.0	-5.0	12.6	4.1
Imports	-0.3	-0.3	-3.4	7.2	2.3	-0.6	-1.3	-11.4	8.4	4.4
Change in:										
Public Debt (% GDP)	-3.8	-0.7	-44.2	3.0	10.0	-3.9	0.6	-86.0	32.7	20.0
External Debt (% GDP)	-2.5	-0.7	-46.8	8.5	10.4	4.1	0.9	-36.5	105.0	24.1
D. PRGT										
Growth	5.6	5.5	1.5	16.7	2.2	5.1	5.2	-3.2	14.0	2.6
Fiscal Adjustment (% GDP)	0.0	0.1	-9.8	1.9	1.6	-0.3	0.1	-33.0	6.1	4.9
Revenue	0.0	0.1	-4.2	3.0	1.1	-0.1	0.2	-11.4	4.2	2.3
Primary Expenditure	0.0	-0.2	-2.4	5.6	1.3	0.2	0.0	-5.8	21.6	3.5
CA Adjustment (% GDP)	0.0	0.4	-10.7	6.3	2.3	0.3	0.3	-15.7	9.7	3.8
Exports	0.5	0.3	-4.2	16.9	2.6	0.2	0.2	-5.6	9.4	2.5
Imports	-0.4	-0.7	-10.2	15.9	3.7	-0.7	-0.6	-17.0	14.1	4.6
Change in:										
Public Debt (% GDP)	-0.7	0.1	-30.7	9.4	5.1	0.4	0.7	-49.6	16.7	7.6
External Debt (% GDP)	0.0	0.4	-26.8	13.5	4.6	0.6	0.6	-52.1	17.9	7.9

Sources: WEO database; IEO staff calculations.
Note: All figures are annual averages over the program period. For consistent comparison, calculations are undertaken for the same sample of programs for which both initial projections and outcomes are available.

Table AVI.2. Results of GRA Growth Regressions: Updated Program Projections

	GRA							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Growth	Growth	Growth	Growth	Growth	Growth	Growth	Growth
ΔPBY	-0.245**	-0.283***	-0.35**	0.256				
	(0.096)	(0.1)	(0.145)	(0.363)				
ΔGREVY					-0.284	-0.504**	-0.431**	0.68
					(0.185)	(0.248)	(0.219)	(0.471)
ΔGEXPY					0.179	0.085	0.117	0.01
					(0.122)	(0.144)	(0.149)	(0.351)
ΔCABY	-0.108	-0.102	0.041	-0.035	-0.079	-0.078	-0.005	-0.029
	(0.133)	(0.133)	(0.088)	(0.112)	(0.104)	(0.079)	(0.09)	(0.132)
TP_GR	0.87***	0.84***	0.96***	0.784**	0.871***	0.893***	0.886***	0.892**
	(0.206)	(0.224)	(0.258)	(0.35)	(0.205)	(0.209)	(0.235)	(0.368)
ΔTOT	0.001	0.002	0.02	0.1	0.01	0.032	0.046	0.09
	(0.051)	(0.047)	(0.053)	(0.14)	(0.051)	(0.048)	(0.057)	(0.151)
Growth (T-1)	0.045	0.052	0.067	0.229	0.082	0.095**	0.078	0.25
	(0.072)	(0.069)	(0.076)	(0.204)	(0.069)	(0.048)	(0.078)	(0.199)
EDY (T-1)	-0.002	-0.001	0.003	-0.005	0.001	0.002	0.007	0.011
	(0.012)	(0.009)	(0.016)	(0.029)	(0.012)	(0.01)	(0.015)	(0.03)
GEI (T-1)	-0.767	-0.242	-0.653	0.193	-1.296**	-1.068*	-0.812	0.204
	(0.659)	(0.81)	(0.694)	(2.053)	(0.578)	(0.614)	(0.663)	(1.942)
EA programs	-0.69	-0.593	-0.777	1.257	-0.66	-0.687	-0.795	1.225
	(0.492)	(0.518)	(0.55)	(1.288)	(0.49)	(0.517)	(0.556)	(1.267)
Crisis programs	0.399	1.52***	0.728	-1.946	0.387	0.93	0.856	-2.083
	(0.574)	(0.446)	(0.633)	(1.829)	(0.552)	(0.661)	(0.648)	(1.863)
Small states	1.195	1.197	1.249	5.083*	2**	1.657*	1.206	7.021**
	(0.839)	(0.967)	(0.799)	(2.755)	(0.848)	(0.874)	(0.976)	(3.239)
PDY (T-1)		-0.012**				-0.015**		
		(0.005)				(0.007)		
ΔPUBINVY			0.126				0.27*	
			(0.174)				(0.156)	
ΔSOCIALY				0.261				0.355
				(1.124)				(1.113)
Constant	0.596	1.652	0.208	0.03	0.163	1.183	-0.061	-0.963
	(1.118)	(1.019)	(1.521)	(2.203)	(1.15)	(1.133)	(1.414)	(2.364)
N	78	61	62	31	76	58	62	31
R^2	0.476	0.527	0.508	0.554	0.509	0.619	0.508	0.562
SE (error term)	1.782	1.514	1.767	2.107	1.725	1.335	1.785	2.163

Source: IEO staff estimates.
Notes: Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table AVI.3. Results of PRGT Growth Regressions: Updated Program Projections

	PRGT							
	(1) Growth	(2) Growth	(3) Growth	(4) Growth	(5) Growth	(6) Growth	(7) Growth	(8) Growth
ΔPBY	-0.261	-0.317*	-0.259	-0.071				
	(0.183)	(0.179)	(0.219)	(0.209)				
ΔGREVY					-0.288	-0.414*	-0.342	-0.033
					(0.231)	(0.237)	(0.256)	(0.222)
ΔGEXPY					0.198	0.113	-0.146	-0.015
					(0.167)	(0.175)	(0.18)	(0.221)
ΔCABY	0.382	0.406	0.402*	0.433*	0.38	0.398*	0.38*	0.44*
	(0.236)	(0.249)	(0.242)	(0.252)	(0.232)	(0.232)	(0.221)	(0.244)
TP_GR	0.069	-0.052	0.000	0.127	0.042	-0.068	-0.007	0.095
	(0.117)	(0.151)	(0.161)	(0.195)	(0.139)	(0.171)	(0.146)	(0.225)
ΔTOT	-0.038	-0.03	-0.027	-0.039	-0.037	-0.025	-0.035	-0.014
	(0.035)	(0.04)	(0.033)	(0.051)	(0.04)	(0.042)	(0.037)	(0.058)
Growth(T-1)	0.114**	0.096*	0.129**	0.095*	0.111**	0.084	0.151***	0.073
	(0.05)	(0.053)	(0.052)	(0.05)	(0.051)	(0.056)	(0.051)	(0.046)
EDY(T-1)	0.013	0.007	0.015	0.007	0.013	0.005	0.012	-0.001
	(0.009)	(0.011)	(0.011)	(0.014)	(0.01)	(0.011)	(0.01)	(0.013)
GEI(T-1)	1.172**	1.25**	0.597	0.859	1.05*	0.892	0.352	0.057
	(0.544)	(0.597)	(0.52)	(0.868)	(0.581)	(0.596)	(0.538)	(0.964)
Small states	-2.525***	-2.576***	-3.125***	-2.040	-2.473***	-2.568***	-3.045***	-1.587
	(0.708)	(0.795)	(1.07)	(1.269)	(0.712)	(0.862)	(1.042)	(1.372)
Fragile states	1.463*	1.552*	1.295	2.321*	1.592*	1.655*	1.29	2.359*
	(0.867)	(0.894)	(0.862)	(1.275)	(0.954)	(0.897)	(0.864)	(1.216)
PDY(T-1)		0.005				0.001		
		(0.015)				(0.016)		
ΔPUBINVY			-0.032				0.336*	
			(0.152)				(0.176)	
ΔSOCIALY				0.293				0.471
				(0.267)				(0.342)
Constant	5.227***	5.792***	5.228***	5.264***	5.269***	5.87***	4.853***	5.388***
	(0.743)	(0.99)	(0.827)	(0.981)	(0.788)	(1.081)	(0.815)	(1.007)
N	204	187	175	107	196	179	175	100
R^2	0.315	0.331	0.367	0.407	0.32	0.357	0.392	0.441
SE (error term)	3.33	3.386	3.3	3.917	3.382	3.393	3.245	3.942

Source: IEO staff estimates.
Notes: Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table AVI.4. Post-Program Potential Growth Regressions (Public Debt Reduction as Stabilization Outcome)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	PTGB(3)	PTGB(3)	PTGB(3)	PTGB(3)	PTGB(3)	PTGB(3)	PTGB(3)	PTGB(3)	PTGB(3)
SCI	-0.071***	-0.072***	-0.068**	-0.059**	-0.069***	-0.057**	-0.07***	-0.075***	-0.078***
	(0.024)	(0.024)	(0.028)	(0.025)	(0.026)	(0.025)	(0.023)	(0.027)	(0.026)
ASCD*SCI	0.377***	0.381***	0.351***	0.249**	0.372***	0.355***	0.374***	0.379***	0.369***
	(0.088)	(0.09)	(0.096)	(0.107)	(0.089)	(0.08)	(0.093)	(0.087)	(0.085)
ASCG*SCI	0.137	0.161*	0.168*	0.191*	0.135	0.075	0.137	0.151	0.176*
	(0.092)	(0.094)	(0.101)	(0.102)	(0.094)	(0.101)	(0.092)	(0.095)	(0.1)
ΔPDY	-0.023**	-0.027***	-0.022**	-0.019	-0.023**	-0.03***	-0.023**	-0.023**	-0.021*
	(0.01)	(0.01)	(0.011)	(0.017)	(0.01)	(0.01)	(0.01)	(0.01)	(0.011)
Debt operations	-1.083*	-1.028*	-1.195*	-0.098	-1.11*	-1.413**	-1.114*	-1.087**	-1.197**
	(0.553)	(0.54)	(0.694)	(0.312)	(0.575)	(0.638)	(0.673)	(0.543)	(0.495)
PRGT programs	1.359**	1.368**	1.25**	1.879**	1.359**	1.587**	1.371**	1.259*	1.007*
	(0.622)	(0.621)	(0.618)	(0.878)	(0.611)	(0.639)	(0.603)	(0.655)	(0.6)
Off-track programs	-0.418	-0.466	-0.372	-0.499	-0.416	-0.251	-0.417	-0.466	-0.597
	(0.415)	(0.423)	(0.443)	(0.52)	(0.414)	(0.413)	(0.389)	(0.462)	(0.442)
Fragile states	-1.353**	-1.424**	-1.429**	-1.225	-1.352**	-1.066*	-1.366*	-1.355**	-1.3**
	(0.631)	(0.641)	(0.616)	(0.836)	(0.619)	(0.591)	(0.717)	(0.628)	(0.617)
Δπ	-0.019								
	(0.09)								
ΔGREVY		-0.048							
		(0.04)							
ΔPUBINVY			0.063						
			(0.04)						
ΔSOCIALY				0.178*					
				(0.097)					
ΔREER					-0.004				
					(0.02)				
GEI						0.525			
						(0.386)			
Small states							0.013		
							(0.87)		
EA programs								-0.204	
								(0.49)	
Crisis programs									-0.667
									(0.467)
Constant	-0.342	-0.38	-0.305	-0.51	-0.355	-0.37	-0.351	-0.2	0.124
	(0.44)	(0.434)	(0.453)	(0.541)	(0.432)	(0.432)	(0.413)	(0.675)	(0.578)
N	52	52	46	31	52	52	52	52	52
R^2	0.397	0.409	0.432	0.514	0.397	0.42	0.396	0.398	0.411
SE (error term)	1.394	1.379	1.407	1.238	1.394	1.367	1.395	1.393	1.378

Source: IEO staff estimates.
Note: Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table AVI.5. Post-Program Potential Growth Regressions (External Debt Reduction as Stabilization Outcome)

	(1) PTGB (3)	(2) PTGB (3)	(3) PTGB (3)	(4) PTGB (3)	(5) PTGB (3)	(6) PTGB (3)	(7) PTGB (3)	(8) PTGB (3)	(9) PTGB (3)
SCI	-0.063**	-0.063**	-0.059**	-0.056**	-0.06**	-0.058**	-0.061**	-0.065**	-0.071**
	(0.025)	(0.026)	(0.028)	(0.027)	(0.028)	(0.024)	(0.025)	(0.029)	(0.028)
ASCD*SCI	0.314***	0.319***	0.29***	0.257**	0.309***	0.306***	0.323***	0.317***	0.312***
	(0.097)	(0.097)	(0.099)	(0.114)	(0.095)	(0.09)	(0.1)	(0.093)	(0.091)
ASCG*SCI	0.081	0.1	0.114	0.145	0.077	0.062	0.082	0.093	0.129
	(0.104)	(0.107)	(0.112)	(0.112)	(0.108)	(0.108)	(0.106)	(0.112)	(0.112)
ΔEDY	-0.017**	-0.018**	-0.017*	-0.01	-0.017**	-0.017**	-0.017**	-0.017**	-0.014
	(0.008)	(0.008)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)
Debt operations	-0.257	-0.128	-0.295	0.048	-0.274	-0.309	-0.515	-0.268	-0.465
	(0.306)	(0.408)	(0.403)	(0.345)	(0.323)	(0.337)	(0.547)	(0.315)	(0.319)
PRGT programs	1.224*	1.242**	1.218*	1.902**	1.208*	1.299**	1.224*	1.154*	0.862
	(0.628)	(0.631)	(0.629)	(0.902)	(0.63)	(0.646)	(0.634)	(0.642)	(0.63)
Off-track programs	-0.792*	-0.828*	-0.712	-0.457	-0.787*	-0.753*	-0.688*	-0.825*	-0.965**
	(0.441)	(0.454)	(0.493)	(0.527)	(0.441)	(0.421)	(0.373)	(0.487)	(0.484)
Fragile states	-0.963	-1.015	-1.028	-1.163	-0.952	-0.861	-1.086	-0.963	-0.906
	(0.69)	(0.696)	(0.714)	(0.855)	(0.69)	(0.659)	(0.811)	(0.69)	(0.681)
Δπ	-0.011								
	(0.085)								
ΔGREVY		-0.038							
		(0.04)							
ΔPUBINVY			0.069*						
			(0.039)						
ΔSOCIALY				0.173**					
				(0.083)					
ΔREER					-0.006				
					(0.022)				
GEI						0.163			
						(0.35)			
Small states							0.404		
							(0.901)		
EA programs								-0.142	
								(0.494)	
Crisis programs									-0.674
									(0.47)
Constant	0.006	-0.038	-0.078	-0.617	-0.006	0	-0.095	0.104	0.472
	(0.421)	(0.419)	(0.464)	(0.505)	(0.42)	(0.421)	(0.372)	(0.647)	(0.607)
N	54	54	47	31	54	54	54	54	54
R^2	0.346	0.355	0.38	0.514	0.347	0.349	0.351	0.347	0.359
SE (error term)	1.468	1.459	1.504	1.239	1.467	1.466	1.463	1.468	1.454

Source: IEO staff estimates.
Note: Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

REFERENCES

Asonuma, M., and others, 2019, "Costs of Sovereign Defaults: Restructuring Strategies, Bank Distress and the Capital Inflow-Credit Channel," IMF Working Paper No. WP/19/69 (Washington: International Monetary Fund).

Atoyan, Ruben, and Patrick Conway, 2006, "Evaluating the Impact of IMF programs: A Comparison of Matching and Instrumental-variable Estimators," *The Review of International Organizations*, 1(2), pp. 99–124.

Atsebi, J., J. Combes, and A. Minea, 2019, "The Trade Costs of Financial Crises", *Études et Documents, n°2, CERDI*.

Atsebi, J., and J. Wojnilower, 2021, "Initiating Growth Surges: The Role of IMF-Supported Programs," IEO Background Paper No. BP/21-01/02 (Washington: International Monetary Fund).

Bal Gunduz, Yasemin, 2016, "Economic Impact of Short-term IMF Engagement in Low-Income Countries," *World Development*, Vol 87.

Barro, Robert, and Jong-Wha Lee, 2005, IMF programs: Who is Chosen and What Are the Effects? *Journal of Monetary Economics*, 52(7), pp. 1245–1269.

Bas, M., and Stone, R., 2014, "Adverse Selection and Growth under IMF Programs," *The Review of International Organizations*, 9(1), 00, pp. 1–28.

Batini, N., L. Eyraud, L. Forni, and A. Weber, 2014, "Fiscal Multipliers: Size, Determinants, and Use in Macroeconomic Projections," IMF Technical Notes and Manuals No. 2014/04 (Washington: International Monetary Fund).

Bird, G., and Rowlands, D., 2017, "The Effect of IMF Programs on Economic Growth in Low-income Countries: An Empirical Analysis," *The Journal of Development Studies*, 53(12), pp. 2179–96.

Blanchard, Olivier and Daniel Leigh, 2013, "Growth Forecast Errors and Fiscal Multipliers," *American Economic Review: Papers and Proceedings*, 103(3), pp. 117-120

Bohn, H., 1998, "The Behavior of US Public Debt and Deficits," *Quarterly Journal of Economics*, 113(3), pp. 949-963.

_____, 2008, "The Sustainability of Fiscal Policy in the United States," in *Sustainability of Public Debt*, edited by R. Neck and Jan-Egbert Sturm, pp.15-49 (Cambridge: MIT Press).

Bordo, Michael D., and Anna J. Schwartz., 2000, "Measuring Real Economic Effects of Bailouts: Historical Perspectives on How Countries in Financial Distress Have Fared with and without Bailouts," *Carnegie-Rochester Conference Series on Public Policy*, vol. 53, no. 1, pp. 81-167.

Butkiewicz, James L., and Halit Yanikkaya, 2005, "The Effects of IMF and World Bank Lending on Long-run Economic Growth: An Empirical Analysis," *World Development*, 33(3), pp.371–91.

Conway, P., 2009, "Participation in IMF programs and Income Inequality," in *Finance Development and the IMF* (edited by James Boughton and Dominico Lombardi).

Dreher, A., 2006, "IMF and Economic Growth: The Effects of Programs, Loans and Compliance with Conditionality," *World Development*, 34(5), pp. 769–788.

Dicks-Mireaux, Louis, Mauro Mecagni, and Susan Schadler, 2000, "Evaluating the Effect of IMF Lending to Low Income Countries," *Journal of Development Economics*, 61(2), pp. 495-526.

Easterly, W., 2005, "What Did Structural Adjustment Adjust? The Association of Policies and Growth with Repeated IMF and World Bank Adjustment Loans," *Journal of Development Economics*, 76, pp.1–22.

Eichengreen, B., P. Gupta, and A. Mody, 2008, "Sudden Stops and IMF-supported Programs," in *Financial Markets Volatility and Performance in Emerging Markets*, pp. 219–266 (Chicago: University of Chicago Press).

Evrensel, Ayşe Y., 2002, "Effectiveness of IMF-supported Stabilization Programs in Developing Countries," *Journal of International Money and Finance*, 21(5), pp. 565-587.

Garuda, G., 2000, "The Distributional Effects of IMF Programs: A Cross-country Analysis," *World Development*, 28(6), pp. 1031–1051.

Ghosh, A., J. Kim, E. Mendoza, J. Ostry and M. Qureshi, 2013, "Fiscal Fatigue, Fiscal Space and Debt Sustainability in Advanced Economies," *Economic Journal,* Vol. 123, No. 566, pp. F4–F30.

Glynn, Adam N., and Kevin M. Quinn, 2010, "An Introduction to the Augmented Inverse Propensity Weighted Estimator," *Political Analysis*,18(1), pp. 36–56.

Gupta, Sanjeev, 2021, "Fiscal Adjustment and Growth in IMF-supported Programs," IEO Background Paper No. BP/21-01/03 (Washington: International Monetary Fund).

Hardoy, I., 2003, "Effect of IMF Programs on Growth: A Reappraisal Using the Method of Matching," Institute for Social Research Paper, No. 2003-040.

Hutchison, Michael M., 2003, "A Cure Worse than the Disease? Currency Crises and the Output Costs of IMF-supported Stabilization Programs," In M.P. Dooley & F. A. Jeffrey (eds.), *Managing Currency Crises in Emerging Markets*, pp. 321–59 (Chicago: University of Chicago Press).

Hutchison, Michael, 2004, "Selection Bias and the Output Costs of IMF Programs," Working Paper Series 04–15, Economic Policy Research Unit, Department of Economics (Copenhagen, Denmark).

Hutchinson, M., and I. Noy, 2003, "Macroeconomic Effects of IMF-sponsored Programs in Latin America: Output Costs, Program Recidivism and the Vicious Cycle of Failed Stabilization," *Journal of International Money and Finance*, 22, pp. 991–1014.

Imai, K. and M. Ratkovic, 2014, "Covariate Balancing Propensity Score," *Journal of the Royal Statistical Society*, Series B: Statistical Methodology, pp. 243-263.

Independent Evaluation Office of the International Monetary Fund (IEO), 2002, *Evaluation of Prolonged Use of IMF Resources* (Washington: International Monetary Fund).

_____, 2003, *Fiscal Adjustment in IMF-supported Programs* (Washington: International Monetary Fund).

International Monetary Fund (IMF), 2015, "Review of Crisis Programs," November (Washington).

_____, 2017, "Regional Economic Outlook: Sub-Saharan Africa," October, (Washington).

_____, 2019a, "2018 Review of Program Design and Conditionality," IMF Policy Paper (Washington).

_____, 2019b, "2018 Review of Program Design and Conditionality—Supplementary Information," (Washington).

_____, 2019c, "Reigniting Growth in Low-income and Emerging market Economies: What Role Can Structural Reforms Play?" *World Economic Outlook*, October, Chapter 3 (Washington).

Ismail, K., R. Perrelli and J. Yang, 2020, "Optimism Bias in Growth Forecasts—The Role of Planned Policy Adjustments," IMF Working Paper WP/20/229 (Washington: International Monetary Fund).

Jordà, Ò., 2005, "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95(1), 161-182.

Jordà, Ò., and Taylor, A., 2016, "The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy," *Economic Journal*, 126, 219–255.

Kim, Jung, and Kwang Y. Lee, 2021, "Structural Conditions, Structural Reforms and Growth in IMF-Supported Programs: 2008-19," IEO Background Paper No. BP/21-01/04 (Washington: International Monetary Fund).

Kuvshinov, D., and Zimmermann, K., 2019, "Sovereigns Going Bust: Estimating the Cost of Default," *European Economic Review*, 119, pp. 1-21.

Lee, Jong-Wha, and Changyong Rhee, 2002, "Macroeconomic Impacts of the Korean Financial Crisis: Comparison with the Cross-country Patterns," *The World Economy*, 25(4), pp. 539-562.

Lunceford, J., and Davidian, M., 2004, "Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study," *Statistics in Medicine*, 23, pp. 2937–2960.

Marchesi, Silvia, and Emanuela Sirtori, 2011, "Is Two Better Than One? The Effects of IMF and World Bank Interaction on Growth," *Review of International Organizations*, Vol. 6, No. 3, pp. 287–306.

Mendoza, E. G., and J. D. Ostry, 2008, "International Evidence on Fiscal Solvency: Is Fiscal Policy "Responsible"?" *Journal of Monetary Economics*, 55(6), pp. 1081-1093.

Mussa, M., and M. Savastano, 1999, "The IMF Approach to Economic Stabilization," IMF Working Paper No. 99/104 (Washington: International Monetary Fund).

Oberdabernig, D., 2013, "Revisiting the Effects of IMF Programs on Poverty and Inequality," *World Development*, 46 (2013), pp. 113–142.

Ostry, J., A. Ghosh, J. Kim and M. Qureshi, 2010, "Fiscal Space," IMF Staff Position Note, 10/11 (Washington: International Monetary Fund).

Przeworski, A., and J. Vreeland, 2000, "The Effect of IMF Programs on Economic Growth," *Journal of Development Economics*, 62 (2), pp. 385–421.

Rosenbaum, P. R., and Rubin, D. B., 1983, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70(1), pp. 41-55.

Rubin, D., 2002, "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation," *Health Services and Outcomes Research Methodology*, 2, pp. 169–188.

Steinwand, Martin, and Randall Stone, 2008, "The International Monetary Fund: A Review of the Recent Evidence," *Review of International Organizations*, 3(2), pp. 123–49.

Vreeland, James Raymond, 2003, *The IMF and Economic Development* (Cambridge: Cambridge University Press).

Wooldridge, J., 2007, "Inverse Probability Weighted Estimation for General Missing Data Problems," *Journal of Econometrics*, 141, pp. 1281–1301.