

21. Despite this heightened attention to growth, the 2018 ROC found evidence that growth outcomes tended to fall short of growth projections even as programs were generally quite successful in solving members' BOP problems in the period covered (2011–17). Its analysis suggested that growth optimism during this period was systematically related to an underestimation of the impact of adjustment on growth. It raised concern that, in the program context, growth optimism could trigger adjustment fatigue and undermine debt sustainability and ultimately program success. The 2018 ROC recommended increased scrutiny of the realism of program baselines and strengthened analysis of the growth impact of program policies. To follow up, the staff is now revising the 2014 OGNC, although this work has been delayed by the heavy work demands of responding to the COVID-19 pandemic.

III. GROWTH AND ADJUSTMENT OUTCOMES OF IMF-SUPPORTED PROGRAMS

22. This chapter provides an overview of growth and adjustment outcomes of IMF-supported programs over the evaluation period, looking at outcomes both during the program and afterwards, based on a range of empirical metrics.⁴

A. Time Pattern of Growth and Adjustment Outcomes

23. In GRA programs, growth outcomes typically exhibited 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 more modest acceleration afterwards (Figure 2). Notable is the wide range of growth outcomes for year T as indicated by the interquartile range in shade. About 41 percent of GRA programs in the sample experienced real GDP contraction (i.e., negative growth) in the first

⁴ This chapter draws on Kim and others (2021) and country case studies prepared for the evaluation.

year (T) of the program, two-thirds of which are accounted for by programs for countries in the context of the GFC and the euro area crisis (“crisis programs”) for which the U-shaped pattern in growth trajectories is particularly pronounced (Figure 3).⁵ Growth outcomes of other programs (which include some programs in response to home-grown BOP crises) were much steadier and show relatively small cross-country variation as indicated by the relatively narrow interquartile range.

24. Consistent with the 2018 ROC, growth outcomes consistently underperformed growth projections, indicating optimism bias embedded in initial program projections. This bias was particularly pronounced in the first year of GRA programs (median bias of 1.5 percentage points) but is also visible in later years (median bias averaging 1.1 percentage points in years T+3 to T+5).⁶ The first-year optimism bias is particularly related to the experience of crisis programs.

25. In PRGT programs, there was a less marked pattern in the trajectory of growth outcomes. In the median program, an initial modest recovery at T was followed by a steady decline in growth until T+3 before leveling off (see Figure 3), in contrast to the steady recovery until T+3 shown in growth projections.

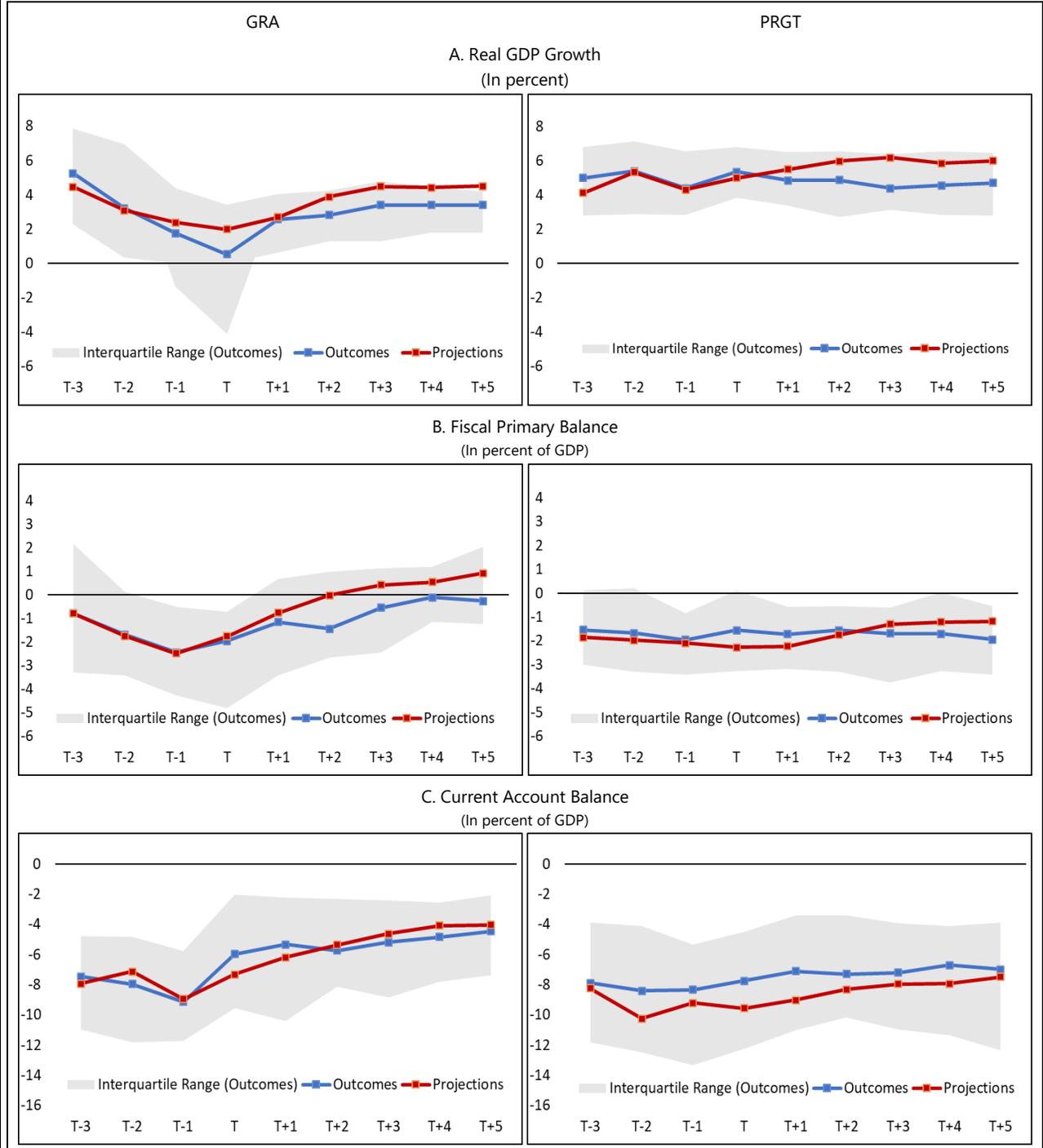
26. As in GRA programs, growth outcomes under PRGT programs generally fell short of projections but with a different pattern. The median outturn was close to projection in the first program year, but fell increasingly short in subsequent years, with the gap reaching 1.8 percent by year T+3.

27. Like growth trajectories, both fiscal and current account (CA) balances exhibited a U-shaped pattern in GRA programs, but the trough was in year T–1 in the case of the CA balance (see Figure 2). Again, the U-shaped pattern observed for fiscal and CA balances was primarily driven by crisis programs. Such a pattern was far less visible in PRGT programs where the trajectories of fiscal and CA outcomes were quite stable over time. GRA programs showed on average smaller fiscal and CA deficits in outcomes and projections than PRGT programs.

⁵ Crisis programs refer to GRA programs arranged in response to a global or major regional crisis. Specifically, crisis programs include GRA programs approved during 2008–09 in response to the GFC (18 programs in total) and five Eurozone programs arranged in response to the euro area crisis (see Table A1 in Annex I for further details). Some other programs have also taken place in the context of BOP crises (e.g., Ukraine 2014 and 2015) but where the source has been internal imbalance rather than an exogenous shock. Their experience has typically followed a similar pattern of adverse growth outcomes.

⁶ This metric does not distinguish between the later years of a multi-year program and the years after the program has been concluded. Thus, the empirical results are not always fully consistent with those discussed in the section “Growth and Adjustment Outcomes Relative to Projections,” which are based on the data for program periods only.

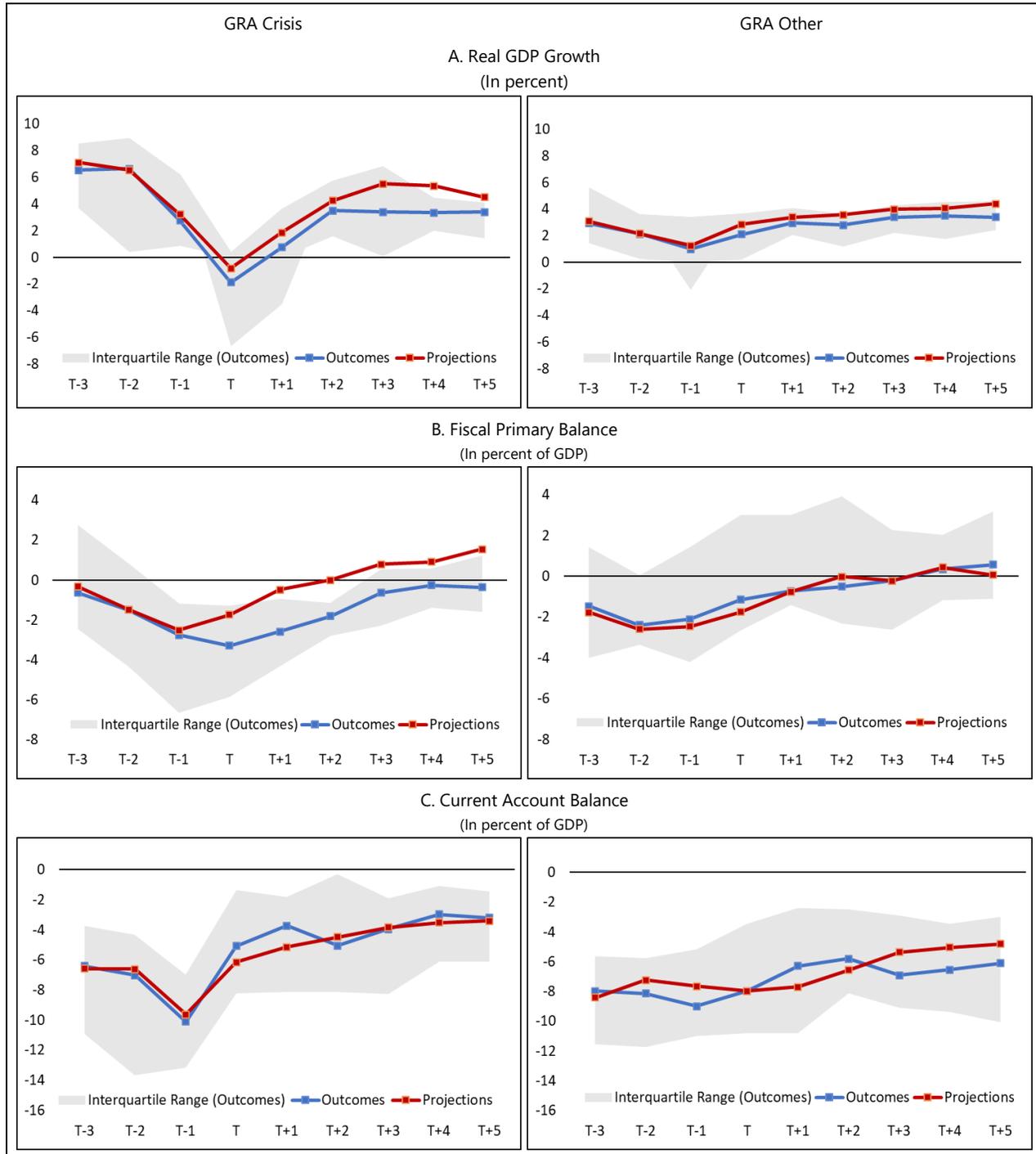
**Figure 2. Growth and Adjustment Trajectories: GRA and PRGT Programs
(Cross-country medians)**



Sources: WEO database; IEO staff calculations.

Note: All projections refer to initial program projections made at program approval (T). 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 some 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 3. Growth and Adjustment Trajectories: GRA Crisis and Other GRA Programs
(Cross-country medians)**



Sources: WEO database; IEO staff calculations.

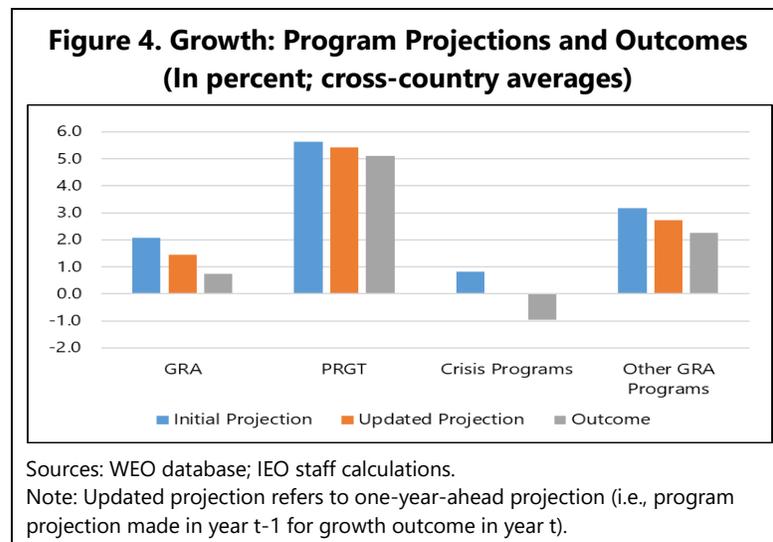
Note: All projections refer to initial program projections made at program approval (T). Outcomes and projections represent cross-country medians. Data availability is not uniform across periods mainly because post-program outcome data are not yet available for recently completed programs. Due to the presence of successor programs for some 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.

28. In GRA programs, median fiscal outcomes were in line with the program in the first and second program years but subsequently underperformed projections by rising margins, particularly in crisis programs, while CA outcomes overperformed initially. In PRGT programs, the median gap between fiscal outcomes and projections was much narrower while CA outcomes consistently outperformed projections. It is notable, however, that CA outcomes varied widely across PRGT programs as indicated by the large interquartile range.

B. Growth and Adjustment Outcomes Relative to Projections

29. A closer look at growth outcomes relative to projections in programs provides further granularity on the growth optimism observed in IMF-supported programs. The analysis in this section is based on the data for program periods only. For consistent comparison between program outcomes and projections, the program sample is limited to 114 programs for which both projection and outcome data are available for one year or longer.⁷

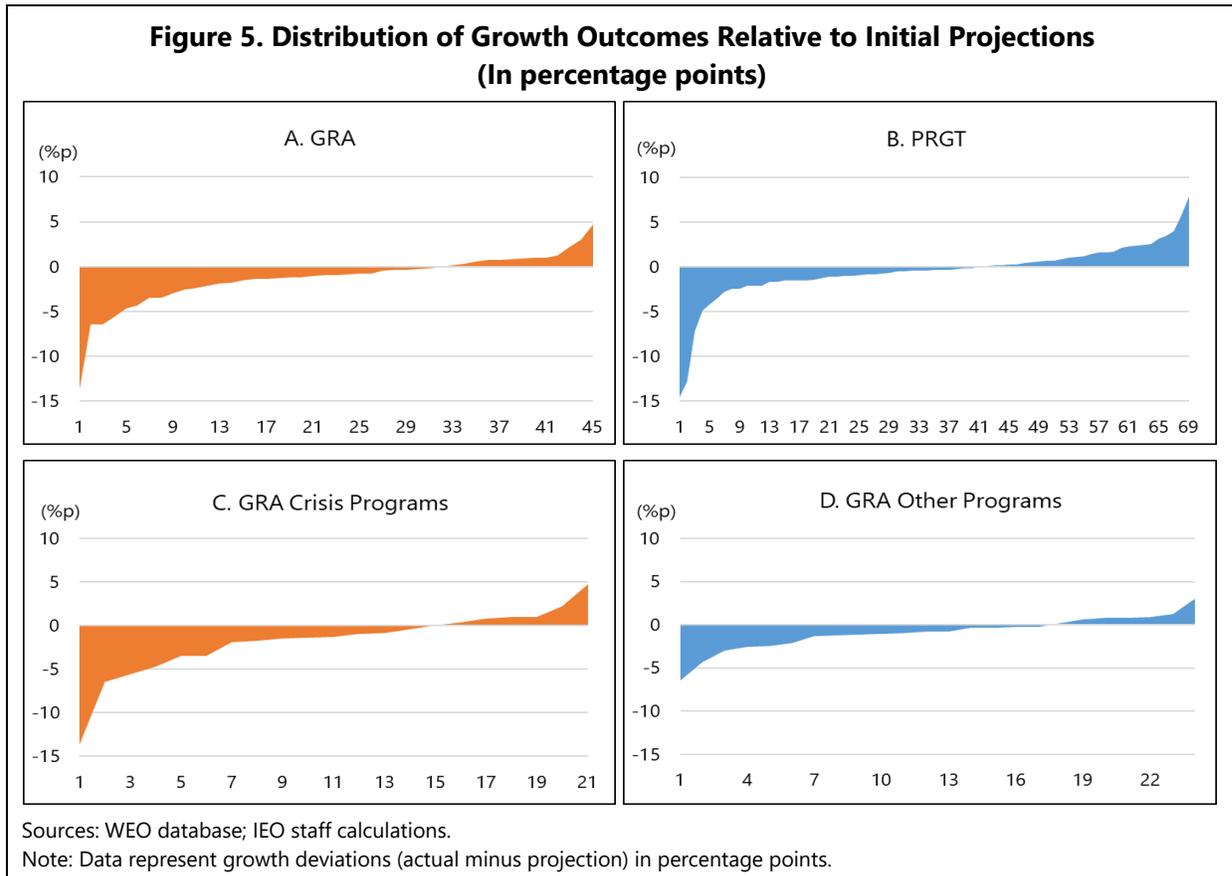
30. **Growth.** Optimism bias in initial program projections averaged 1.3 percentage points in GRA programs—somewhat larger in crisis programs than in other GRA programs—and 0.5 percentage points in PRGT programs (Figure 4).⁸ In both GRA and PRGT programs, growth projections were revised downwards over the program period, generally more so in GRA programs and particularly in crisis programs than in PRGT programs. Thus, updated (one-year-ahead) program projections were typically more realistic than initial projections in both GRA and PRGT programs.



⁷ Some programs in the evaluation sample went quickly off track; as a result, no observations are available for program outcomes under the conventions used to determine program duration for analytical purpose (see Annex I). Comparison is based on annual averages over the program period. See Kim and others (2021) for further technical details.

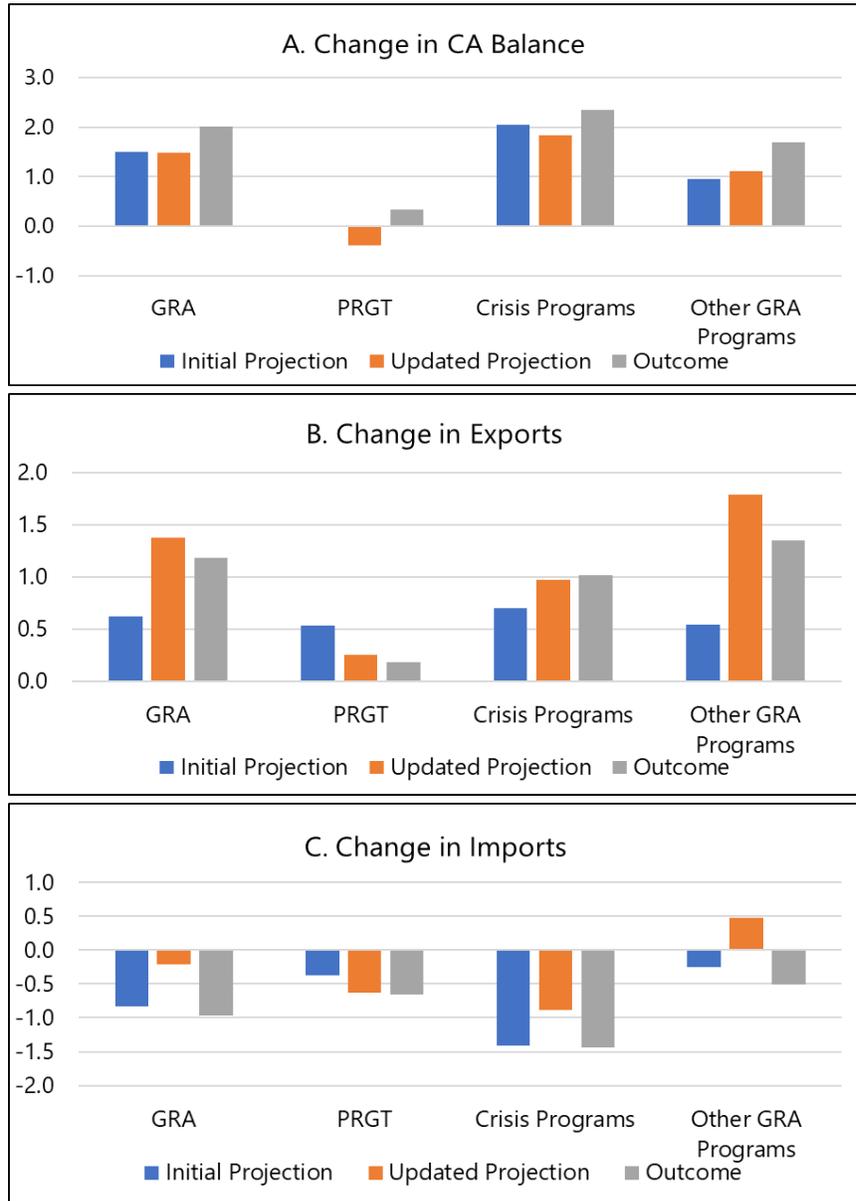
⁸ Baseline projections at program approval assume full program implementation, implying that less than full program implementation could ex post lead to optimism bias. Similarly, ex post data revisions including GDP rebasing could be a source of optimism bias by itself and by affecting modeling errors in program design.

31. Growth shortfalls were widely dispersed across both GRA and PRGT programs (Figure 5). For GRA programs, growth fell short of projections by more than an annual average of ½ percentage points in 58 percent of cases; in 25 percent of cases the growth shortfall was greater than 2.2 percentage points. For PRGT programs, growth shortfalls were larger than ½ percentage points of GDP in 42 percent of cases and larger than 1.5 percentage points in 25 percent of cases.



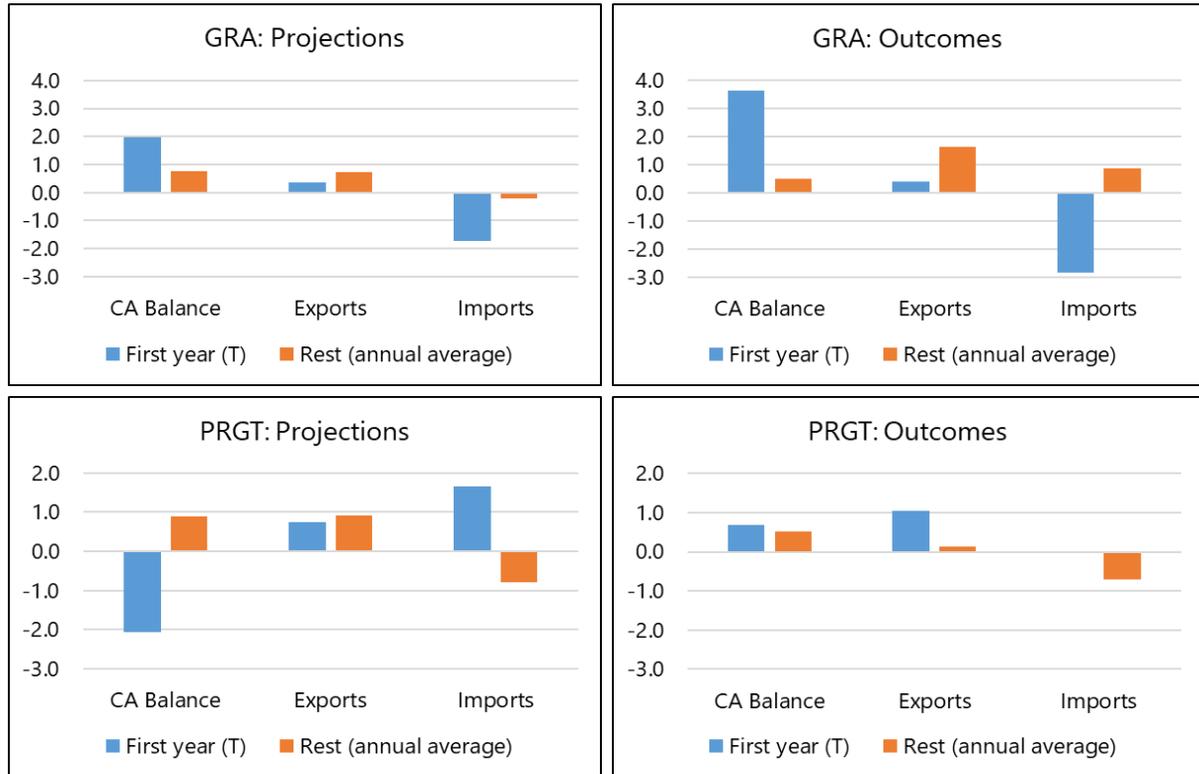
32. **External Adjustment.** In GRA programs, on average actual CA adjustment exceeded modestly (by ½ percentage points) the programmed adjustment (Figure 6). Within GRA programs, both programmed and actual CA adjustments were stronger and relied more on import compressions in crisis programs than in other programs where improved exports played a greater role than import compression. Programmed and actual CA adjustments were both far smaller in PRGT programs but subject to large cross-program variation. Programmed CA adjustment was front loaded in GRA programs but back loaded in PRGT programs (Figure 7). In GRA programs, front loading was even more pronounced in program outcomes, largely driven by import compression. In sharp contrast to program projections, actual CA adjustment in PRGT programs was evenly phased, largely because projected increases in imports in the early phase of the program did not materialize.

**Figure 6. External Adjustment: Program Projections and Outcomes
(In percent of GDP; annual average)**



Sources: WEO database; IEO staff calculations.

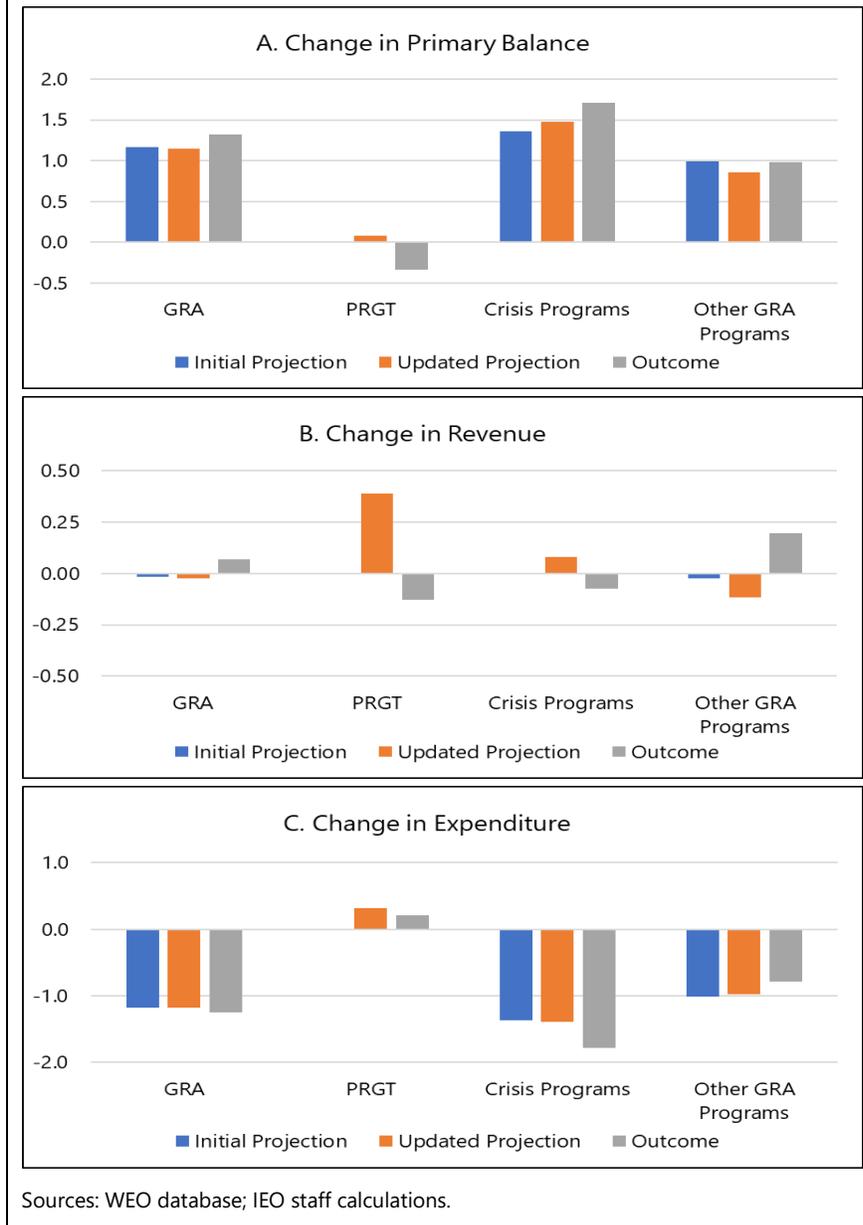
**Figure 7. Phasing of External Adjustment: Initial Projections and Outcomes
(In percent of GDP)**



Sources: WEO database; IEO staff calculations.

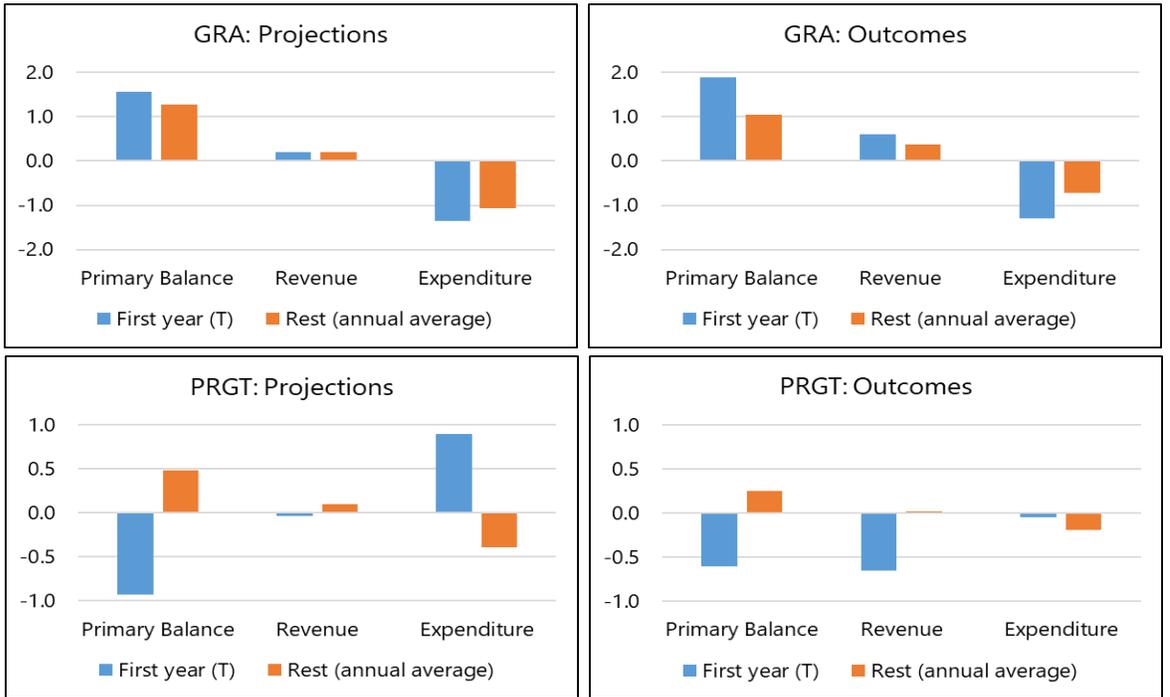
33. **Fiscal adjustment.** In contrast to growth, actual improvement in the primary balance on average were slightly higher than program projections in GRA programs but fell short in PRGT programs (Figure 8). Within GRA programs, differences between fiscal outcomes and projections were on average larger in crisis programs where fiscal outturns were significantly stronger than projected, particularly in expenditure adjustment. Programmed fiscal adjustment was on the order of 1.2 percent of GDP in GRA programs on an annual average basis but small in PRGT programs. Adjustment was dominated by expenditure adjustment in both projections and outcomes in GRA programs, while the composition was more even in PRGT programs. 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 initial fiscal easing in the first year (T) of the program followed by fiscal tightening in the rest of the program period (Figure 9).

Figure 8. Fiscal Adjustment: Program Projections and Outcomes
(In percent of GDP; annual average)



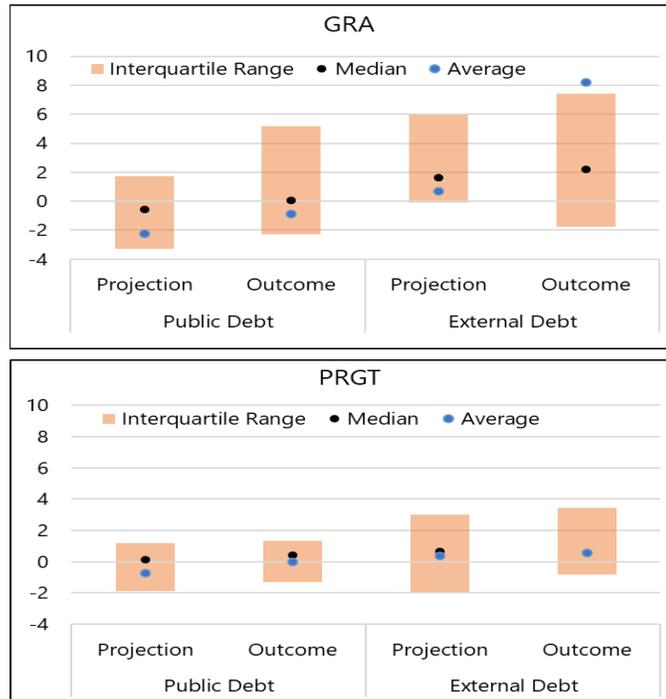
34. **Debt.** The combination of somewhat weaker growth outcomes (particularly in GRA cases) with more modest fiscal consolidation efforts (particularly in PRGT cases) has meant that public debt-to-GDP ratios have tended to rise rather than decline as programmed in both GRA and PRGT programs. Moreover, there is large cross-country variation in debt projections and outcomes especially in GRA programs as indicated by large interquartile ranges (Figure 10).

Figure 9. Phasing of Fiscal Adjustment: Initial Projections and Outcomes (In percent of GDP)



Sources: WEO database; IEO staff calculations.

Figure 10. Distribution of Change in Debt (In percent of GDP)



Sources: WEO database; IEO staff calculations.

C. Growth Outcomes Relative to Benchmark

35. This section seeks to compare growth outcomes during program periods to a historical growth benchmark for each country that seeks to reflect the impact of exogenous changes in the country's growth environment unrelated to the program but does not take into account a country's adjustment needs. This exercise recognizes that IMF-supported programs in the evaluation sample were approved and completed at different times against different cyclical 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.

36. The growth benchmark used in the evaluation is constructed to capture the variation in actual growth explained by external factors as well as country-specific historical trend growth. The benchmark is based on panel regressions linking growth to exogenous factors estimated for 174 countries over the period 1990–2019, including non-program as well as program periods. Growth deviations from the estimated benchmark are then calculated for program periods, and by construction should reflect primarily the influence of domestic factors such as domestic policy adjustments and supply shocks.⁹

37. Applying this approach, average growth deviations from the estimated benchmark ranged widely from –11.2 percent for Ukraine (2008 SBA) to 9.5 percent for Afghanistan (2011 PRGT) (Figure 11). While growth deviations are relatively evenly split between positive and negative values, the GRA sample is populated largely by negative deviations while the opposite is the case for the PRGT sample. As a result, the sample median diverges significantly between GRA (–1.5 percent) and PRGT programs (0.9 percent). Within GRA programs, sample medians also differ significantly between crisis programs (–3.5 percent) and other programs (–0.7 percent).

38. This exercise suggests that there were relatively few cases (12 percent of the full sample) in which the program growth outcome fell significantly below the country's historical norm, mostly associated with crisis programs with large adjustment needs, as well as some later programs with countries like Ukraine facing acute home-grown BOP problems. Overall, positive or negative growth deviations from the benchmark were statistically significantly different from zero (at 10 percent or higher) in 24 out of the 120 programs in total. The distribution of statistically significant deviations is quite uneven between the GRA and PRGT samples—positive deviations are entirely from the PRGT sample while almost all negative deviations are from the GRA sample (Table 1). Within the GRA sample, crisis programs dominate other programs in

⁹ See Kim and others (2021) for detailed discussion of the estimation of growth benchmarks and related empirical findings. The benchmark is not intended to be a counterfactual (e.g., growth outcome that would have prevailed with no Fund engagement).

accounting for negative and significant growth deviations—11 out of 13 negative and significant deviations in the GRA sample are associated with crisis programs.¹⁰

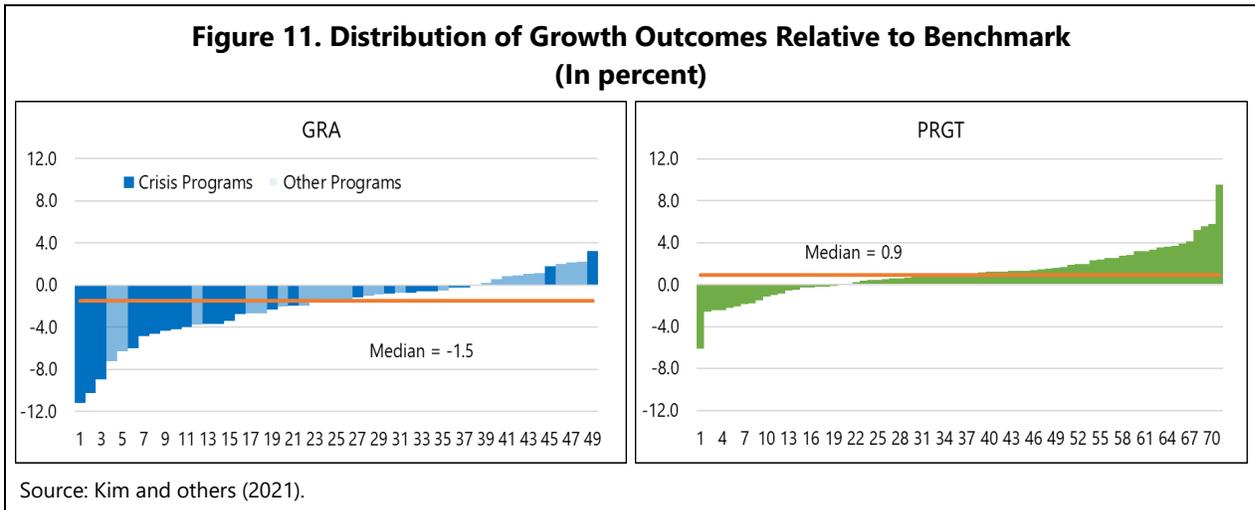


Table 1. Distribution of Growth Deviations by Program Type

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: Kim and others (2021).

Note: Figures in parentheses are the number of programs for which the average growth deviation is statistically significantly different from zero at 10 percent or higher.

39. Country case studies undertaken for the evaluation provide some further insights about the country-specific drivers of growth in cases of statistically significant growth deviations from benchmark. Positive growth deviations were often associated with favorable supply-side factors such as new mines coming on stream and good harvest in Ghana (2009 PRGF), a post-flood rebound in agriculture and buoyant tourism in Grenada (2014 ECF), and a demand stimulus from surge in public investment financed by capital inflows in Senegal (2015 PSI).

40. Negative and significant growth deviations found in case studies were mostly associated with crisis programs and driven by a range of negative demand and supply shocks as well as political factors. Latvia (2008 SBA) and Romania (2009 SBA) were both afflicted by an unwinding of an unsustainable economic boom and severe credit crunch in the aftermath of the GFC, as well

¹⁰ The remaining two GRA programs with negative and significant growth deviations are Ukraine (2014) and Surinam (2016), both of which went off-track. Sierra Leone (2013) is the only PRGT program indicated with a negative and significant growth deviation.

as fiscal consolidation. Mongolia (2009 SBA) was hit hard by a slowdown in investment flows to the mineral and construction sectors and further by strong fiscal consolidation implemented in the early phase of the program. For Ukraine (2008 SBA) which went quickly off track, restricted access to international capital markets after the GFC and a sharp fall in exports led to a sharp recession. Moreover, pre-existing domestic vulnerabilities, reflecting a stalled transition to a market-oriented economy and poor economic governance, weighed on growth. Ukraine (2014 SBA), which also went off track, suffered from political unrest and military conflict in the Eastern region which overwhelmed the government and depressed economic confidence.

D. Sources of Growth Optimism

41. It has long been recognized that the IMF's short-term growth forecasts are subject to optimism bias, particularly outside the advanced economies (Timmermann, 2007; IEO, 2014). The 2018 ROC confirmed optimism bias in growth projections in the program context and sought to identify its origin, following the approach used by Blanchard and Leigh (2013). More specifically, in the full ROC sample, short-run optimism bias was found to be slightly more than 1 percentage points, about one-quarter of which was accounted for by underestimation of the growth impact of fiscal and CA adjustments. Another one-quarter was explained by forecast errors of external conditions.

42. For this evaluation, we investigated growth optimism by undertaking a cross-country analysis of growth forecast errors drawing on the approach used in the 2018 ROC but for a slightly different purpose of assessing whether the contribution of macroeconomic modeling errors to growth optimism differed between program and non-program periods and between GRA and PRGT programs. Regression analysis was undertaken for a panel sample of 75 countries included in the evaluation sample over the period 2009–19.

43. Based on this approach, we found, similar to the analysis of the 2018 ROC, that growth forecast error regressions explain about one-quarter of total sample variation in growth forecast errors, leaving a large unexplained variation. The estimation results reported in Kim and others (2021) suggest that large planned fiscal adjustments were associated on average with smaller optimism bias than average-sized fiscal adjustments. This finding, although at odds with evidence found by the 2018 ROC and other related studies,¹¹ may reflect that confidence effects associated with larger fiscal adjustments helped to offset income effects captured by standard multiplier analysis. Another finding is that macroeconomic modeling errors related to too low fiscal multiplier assumptions (relative to the estimated actual) were a statistically significant

¹¹ For instance, Ismail, Perrelli, and Yang (2020) find for a large panel sample of 170 countries for the period 2003–17 that large fiscal adjustments (one-half standard deviation or more above the sample average) is associated with higher growth optimism in surveillance and non-concessional program forecasts.

source of growth optimism in GRA programs other than crisis programs, although not in crisis or PRGT programs.¹²

44. While macroeconomic modeling errors have played a role, persistent growth optimism across programs seems to be substantially related to other factors. Drawing on case study evidence, an important role seems to have been played by political economy considerations in difficult program negotiations. Authorities have an incentive to provide the public with prospects of a robust payoff from adjustments and reforms to garner needed political support. Fund staff may also have an incentive to agree to unrealistic growth projections, which make it technically easier to close fiscal gaps and reach favorable conclusions about debt sustainability, while hoping to convince authorities to advance difficult adjustment and reforms. Several case studies illustrate how such factors played out in practice. In Latvia (2008), Fund staff anticipated a GDP decline of 6 percent to 8 percent in 2009 given the data already pointing to a sharp recession, but agreed to program a 5 percent decline as the authorities viewed such a forecast as overly pessimistic; the eventual outcome was a 14 percent contraction. In the case of Jamaica, staff noted in interviews with the IEO that medium-term growth forecasts were probably overoptimistic but cautioned that it would have been challenging to get domestic support for a program with even lower medium-term growth projections. In Jordan, Pakistan, and Tunisia, Fund staff underestimated the complexity of the political transition and the impact of intervening political, security-related and regional shocks. At the same time, country officials wanted to show hope to sustain political support for challenging reforms. The consequence was a disconnect between optimistic growth projections and actual outcomes.

E. Assessment

45. Overall, the evidence suggests that IMF-supported programs in the evaluation sample did not demonstrate consistently adverse growth outcomes after accounting for adjustment needs and the external environment. The most negative growth outcomes occurred in the first year of GRA-supported crisis programs. More generally, in the large majority of programs, growth outcomes did not fall significantly short of a historical growth benchmark that corrects for the influence of exogenous external factors and the difference in historical trend growth, but does not take into account adjustment needs. By this metric, PRGT programs appear more successful in achieving growth than GRA programs, which can be attributed in part to the fact that adjustment needs were generally smaller in PRGT programs than in GRA programs. By the same token, the adverse growth outcomes of crisis programs can be attributed to strong policy adjustments needed to address acute BOP pressure, financial fallout from the GFC and associated severe credit crunches and, in some cases, political unrest.

¹² The variance decomposition results reported in Kim and others (2021) show that modeling errors related to fiscal multipliers explain 30 percent of sample variation in growth forecast errors (after country and vintage fixed effects) in GRA programs other than crisis programs while little in crisis and PRGT programs.

46. IMF-supported programs also delivered substantial adjustment in terms of the external current account balance and the fiscal primary balance. Fiscal adjustment was large (on the order of 1 percent of GDP per year) and significantly front loaded in GRA programs, especially in crisis programs where restoring investor confidence early on would likely be key to program success. In these programs, adjustment was achieved almost entirely through spending cuts. In PRGT programs, by contrast, fiscal adjustment was backloaded (i.e., initial easing followed by tightening). Such different magnitude and pattern of fiscal adjustment between GRA and PRGT programs are reflective of the differences in the nature and sources of BOP pressure, market access, adjustment need, debt sustainability concerns, and program objectives.

47. Program projections over the evaluation period were subject to considerable growth optimism bias, reinforcing the findings of the 2018 ROC. Over the programs covered in the evaluation, around one-half experienced an average growth shortfall (relative to initial program projections) during the program period of ½ percentage points or more, while one-fourth had a growth shortfall of over 1½ percentage points. Optimism bias was on average larger in GRA programs (particularly in the first year of crisis programs) than in PRGT programs, although PRGT programs showed rising growth shortfalls in later years. Macroeconomic modeling errors, particularly those arising from unrealistic program assumptions on fiscal multipliers, seem to have been an important source of growth optimism in GRA programs other than crisis programs but less so in crisis or PRGT programs. In GRA crisis programs where fiscal adjustment was far stronger than in other programs (see Figure 6), the seemingly limited role of fiscal modeling errors in accounting for variation in growth forecast errors may be related to positive confidence effects that large and front-loaded fiscal adjustment can entail and help to offset in part negative income effects of fiscal adjustment.

48. Case study evidence suggests that while macroeconomic modeling errors played a role, political economy factors in difficult program negotiations also contributed to growth optimism in program design. Several case studies illustrate pressures on staff and the authorities to agree on excessively sanguine projections, hoping to sustain domestic support but underplaying the risks of subsequent growth disappointments and the challenges of program implementation.

49. Persistent growth optimism raises serious concerns because growth outcomes below program goals tend to contribute to adjustment fatigue, undercut program ownership, and fuel skepticism and rising opposition to reforms. While greater scrutiny of the realism of program projections as recommended by the 2018 ROC could help to reduce growth optimism, it seems even more relevant to consider whether IMF-supported programs can achieve growth outcomes more in line with growth projections by paying greater attention to growth-friendly policies in program design, as assessed in the later chapters of this report.