Quality of Macro-Fiscal Frameworks in Development Policy Operations

IEG Learning Product

June 15, 2015

Contents

Boxes

Tables

Figures

Evaluation Managers	
❖ **Caroline Heider**	Director-General, Evaluation
❖ **Nicholas David York**	Acting Director, Independent Evaluation Group, Public Sector
❖ **Mark Sundberg**	Manager, Independent Evaluation Group, Public Sector
❖ **Zeljko Bogetic**	Task Manager

Abbreviations

AAA	Analytic and Advisory Activities
AFR	Africa
CEMs	Country Economic Memorandum
CPIA	Country Policy and Institutional Assessment
DPL	Development Policy Loan
DPOs	Development Policy Operations
DSA	Debt Sustainability Analysis
ECA	Europe and Central Asia
EP	Economic Policy
ESW	Economic and Sector Work
GDP	Gross Domestic Product
HD	Human Development
IBRD	International Bank for Reconstruction and Development
ICR	Implementation Completion and Results Report
ICRR	Implementation Completion and Results Review
IDA	International Development Agency
IEG	Independent Evaluation Group
IMF	International Monitory Fund
JBS	Joint Budget support
MENA	Middle East and North Africa
MTEF	Medium-Term Expenditure Framework
NLTA	Non Lending Technical Assistance
OPCS	Operations Policy and Country Services
PEFA	Public Expenditure and Financial Assessment
PEMSP	Public Expenditure Management
PER	Public Expenditure Review
PFM	Public Financial Management
PGS	Policy-Based Guarantees
POV	Poverty
PPP	Public Private Partnerships
PREM	Poverty Reduction and Economic Management Network
PRSC	Poverty Reduction Support Credit
PS	Public Sector
PSIA	Poverty and Social Impact Analysis
RETFs	Recipient Executed Trust Funds
SPF	State and Peace Building Fund
SDN	Sustainable Development Network

Acknowledgments

This paper is a Learning Product that was prepared under the general direction and guidance of Mark Sundberg, Manager, Independent Evaluation Group Public Sector (IEGPS).

The database, analysis, and learning product were prepared by the IEG team consisting of Zeljko Bogetic (team leader), Javier Bronfman, Malathi Jayawickrama, Marko Klasnja, Aghassi Mkrtchyan, Moritz Piatti, Marcelo Selowsky, and Jesse Torrence. The team leader would like to express appreciation to Mark Sundberg for substantive guidance and advice in all stages of this study. Peer reviewer comments from Ismail Arslan, Pedro L. Rodriguez, Shahrokh Fardoust, as well as comments and discussions with Linda van Gelder, Ed Mountfield, Manuela Francisco, Peter Moll, and Jasmin Chakeri are gratefully acknowledged.

Individual contributions of team members, who provided building blocks of the complete analysis presented in the paper, are highly appreciated. They are as follows: Moritz Piatti (management, improvements, and revisions of the Development Policy Operations (DPO) and Public Expenditure Review (PER) databases based on the early database developed by Jesse Torrence; detailed portfolio analyses of both DPOs and PERs; authorship of one case study, and working closely with team leader and Marko Klasnja on the statistical analysis); Javier Bronfman (author of several case studies, who also synthesized case studies for this paper on macro-fiscal frameworks), Marko Klasnja (data issues, statistical analysis), Marcelo Selowsky (author of several case studies), Malathi Jayawickrama (author of several case studies, who also prepared a draft synthesis of case studies on public expenditures and DPOs in the accompanying paper on public expenditures), Aghassi Mkrtchyan (author of several case studies), and Jesse Torrence (who developed an early database of DPOs). Team assistance by Aline Dukuze and Yezena Zemene Yimer is gratefully acknowledged.

Overview

This paper is part of a series of Learning Products (LP) that primarily synthesize existing evaluative material with the aim of gaining new insight into the effectiveness and design of development policies and interventions supported by the Bank from an operational perspective. In this series the focus of the LP is on factors that influence design, policy implementation, and performance of Development Policy Operations (DPOs). This particular report aims to assess the quality of the macro-fiscal frameworks in DPOs using a variety of methods (portfolio review, case studies, and statistical analysis) to derive insights, lessons, and good practices that could inform the future design of DPOs.[1]

The three overarching questions in this paper are the following:

What has been the quality of macro-fiscal frameworks and their building blocks?

What factors may be related to the quality of macro-fiscal frameworks and the success of DPOs?

What good practices can be identified that may provide lessons for the future design and implementation of DPO macro-fiscal frameworks?

We approach these questions using an eclectic analytical approach combining:

- *elements of a portfolio review of DPOs* with particular focus on the macro-fiscal frameworks and their links to macro-fiscal policy reforms supported under DPOs,
- *a comparative review based on 15 case studies* (out of the total of 46 DPOs closed in FY11-13 with Implementation Completion Report [ICR] reviews). These provide deeper narratives of the quality and factors that influence the design and implementation of macro frameworks and their contribution to DPO performance. Case studies were selected to include all world regions, programmatic and standalone DPOs, and different sectoral types of DPOs. These case studies unintentionally over-represent DPOs with less than satisfactory ratings to some extent.[2]

[1] For the purpose of this analysis, we adopt the following operational definitions rooted in Operations Policy and Country Services (OPCS) guidance and regulations and operational practice. *Adequacy* refers to the statutory requirement that DPOs must have a macro-fiscal framework that meets certain minimum quality requirements in three dimensions: internal consistency, credibility, and debt sustainability (OPCS, September 2013). *Quality* refers to the degree to which different building blocks of macro-fiscal frameworks—consistency, credibility, and sustainability—can be assessed as having different levels of strength upon careful inspection and analysis. For the purpose of this analysis, *consistency* is defined in terms of completeness of macro-fiscal frameworks, coherence between objectives and measures, and realism. *Credibility* refers to the government's track record with macro-fiscal management in the previous three years as well as the realism of its fiscal program. And *sustainability* refers to completeness and quality of the medium-term fiscal and debt analysis.

[2] The percent of DPOs closed in FY2011-13 that are rated by the IEG as moderately unsatisfactory or lower was 24 percent. By contrast, the percent of all individual DPOs in the 15 case studies (including several DPO series) rated by the IEG as moderately unsatisfactory or lower was 31.5 percent.

- *preliminary statistical analysis* to understand the macroeconomic framework-related correlates of success of DPOs, as well as the correlates of the quality of macro-fiscal framework itself.

This analysis triangulates and synthesizes the evidence to draw the conclusions reported. This includes some thoughts toward improving the design and implementation of macro-fiscal frameworks in future DPOs as well as questions for further research. Results presented here are meant to answer specific questions using the data available, but also raise questions for further research.

The target pool of DPOs and the periodicity under study is the following: first, all DPOs that were closed and having ICR review during FY05-13—a total of 390 DPOs—were included in the database of basic characteristics of DPOs. This pool and related data are used in the preliminary portfolio review and statistical analysis. Second, closer attention was paid to the 46 DPOs that closed and had ICR reviews completed in FY11-13. About a third of these recent DPOs (15 in total) were selected for in-depth, desk-based country case studies to uncover deeper country-level and project-level narratives and elements of good practice (and weakness) in the quality of macro frameworks. A comprehensive Independent Evaluation Group (IEG) DPO database was developed for this purpose and it informed this learning product and other DPO learning products in the series.

The main conclusions follow:

I. Insights from a portfolio analysis focused on macro-fiscal issues

The majority of DPOs (70 percent) were managed by the Poverty Reduction and Economic Management Network (PREM) of the World Bank, which, following the reorganization in July 2013, is now organized into three global practices—Macroeconomics and Fiscal Management (MFM), Governance, and Poverty—and Gender cross-cutting area. This is followed by Human Development (HD)—now represented by Education, Health, and Social Protection and Labor global practices (12 percent). Next is the Sustainable Development Network (SDN), now represented by the global practices Urban, Rural and Resilience, Transport & ICT, Energy & Extractives, Environment and Natural Resources, and Agriculture (10 percent). Within PREM, most DPOs were managed by Economic Policy (EP)—current Macroeconomics and Fiscal Management global practice. The Poverty (POV) and Public Sector (PS) global practices each managed 12 percent of all DPOs. About 80 percent of all DPOs were rated moderately satisfactory or above by IEG.

Programmatic DPOs appear to have performed better in terms of IEG ratings than standalone operations between 2005 and 2008. Thereafter, performance was relatively equal. Most DPOs were multi-sector operations (they could be programmatic or standalone), but single-sector operations seem to be increasing. Multi-sector DPOs have been performing better than single sector DPOs over the years. These independent results are broadly consistent with and corroborate the recent empirical analysis of the correlates of DPO success (Moll, Geli, and Saavedra 2015). The reasons are not explored in this report, but may be related to better design, staff skills, and greater managerial scrutiny and, on the demand side, self-selection bias of countries with stronger policies and institutions.

There has been progress in simplifying policy frameworks of DPOs in terms of the number and complexity of prior actions. The average number of prior actions fell from 16 to nine between 2004 and 2008. There is some evidence suggesting that complicated prior actions with multiple subactions are being replaced by simpler and more concrete formulations. This is in line with OPCS guidance on simplification, reducing the number, and increasing concreteness, relevance, and quality of prior actions. After 2008, the average number has stabilized between eight and 10 (also, see another accompanying Learning Product in this series on results frameworks in DPOs, IEG 2015b).

The thematic content of DPOs reflects their policy focus on mainly public sector governance and macro-fiscal issues. The majority of all DPO prior actions are coded for the broad theme clusters of public sector governance (43.5 percent), now represented by followed by financial and private sector development (18.3 percent), and human development (10.5 percent).

A more granular breakdown of DPO prior actions by specific themes shows that macro-fiscal and public financial management (PFM)-related prior actions make up about 30 percent of all prior actions. Within this macro-fiscal and PFM theme, public expenditure, financial management, and procurement is by far the largest category (23.6 percent), followed by tax policy and administration (4.4 percent), and macroeconomic management (2.9 percent). Other prior actions beyond macro-fiscal and PFM, for example, include administrative and civil service reform (7.7 percent), regulation and competition policy (7.0 percent), and education for all (5.3 percent).

The implementation of an adequate macroeconomic framework was most frequently used as a macro-economic prior action. Some 69 percent of prior actions targeted broad maintenance of sound macroeconomic policy, followed by 27 percent of macro-fiscal actions specifically targeting the budget (e.g., submission of the budget or its amendment) and 19 percent concrete expenditure policy measures.

In terms of the nature of macro-fiscal prior actions, most relate to policy implementation. Prior actions that are more documentary (e.g. the Ministry of Finance has issued a report) were 21 percent. Legislative reform (e.g., preparing, submitting, or passing a bill) accounts for 10 percent of all macro-fiscal actions. Tax policy actions account for only 4.4 percent of macro-fiscal and PFM actions.

Given that many DPOs target fiscal adjustments, the relative infrequency of explicit prior actions on tax policy requires further analysis. This is puzzling in view of considerable structural, development-critical issues in tax policy (e.g., exemptions, tax expenditures, tax base, equity, subnational taxation) in many countries and adjustment programs.

Consistency[3] of the framework was assessed by reviewing macroeconomic frameworks of 174 of the total of 390 DPOs. In this sample, consistency was solid (adequate or better) for the majority of operations. On a 5-point scale (from 1, the lowest, to 5, the highest) the average score in the sample was 3.4. A small number of operations (4 percent) showed significant weaknesses or especially limited treatment of the macro-fiscal framework.

Among Regions, Europe and Central Asia has the largest share of DPOs with a solid consistency (rating 3 and above) and the South Asia Region had the lowest share.

Consistency of macro-fiscal frameworks showed improvement over time. There has been a steady and significant increase in the share of operations scoring three, four, and

[3] For the purpose of this analysis, *consistency* is defined in terms of completeness of macro-fiscal frameworks, coherence between objectives and measures, and realism. Also, within the same analysis, the team assessed the completeness of the debt *sustainability* analysis.

five over the years using this metric. Only in 2012 there appears to be somewhat of a decline, but this may reflect the small sample size.

II. Insights and lessons from case studies

Overall, the macro-fiscal frameworks presented in DPO desk reviews of a number of case studies are broadly adequate. Basic macro frameworks in most cases are consistent, credible, and sustainable as presented in the program document as well as in related evidence reviewed (e.g., IMF reports, detailed debt sustainability analysis). This conclusion is consistent with the aforementioned independent analysis of the consistency of macro-fiscal frameworks, as well as recent research on the correlates of DPO success (Geli, Moll, and Saavedra 2015).

Some good practice DPOs are identified (Peru and Romania series), which are well designed programmatic operations that also integrate well the World Bank's knowledge on public expenditures with the DPO policy framework, as documented in the accompanying paper (IEG learning product on the role and integration of public expenditure knowledge with DPO design, 2015a).

In most cases, the macro-fiscal framework is highly aligned with the IMF and related strategic, lending, and analytic and advisory activities (AAA) programs. This includes IMF program documents or Article IV consultations, as well non-lending technical assistance (NLTA), economic and sector work (ESW), and other AAA relevant work (e.g. a good example is the Peru DPO). (Also, see IMF 2010).

At the same time, some weaknesses in the quality of macro frameworks are identified in a few operations. Three broad types of weaknesses are highlighted: (i) ambitiousness of macro-fiscal frameworks in some standalone operations and links between objectives and fiscal measures, (ii) credibility of the framework in view of track record, political economy factors, treatment of risks, or institutional fiscal rules, and (iii) robustness of the debt sustainability analysis. Some examples are the following:

- In some cases, the targets were too ambitious relative to what the operation supported. In standalone operations with multiple objectives, for example, highly ambitious macro targets are hard to achieve within the time envisaged (e.g., St. Lucia). In other cases, fiscal measures in support of objectives are not sufficiently articulated or left implicit. It was also found that greater attention was sometimes needed to link objectives to fully articulated macro-fiscal measures.
- In terms of the credibility of the macro and fiscal frameworks, a strong track record of fiscal prudence and an existing robust institutional framework, such as fiscal rules, was correlated with better ability to advance reforms while keeping the finances in good shape (e.g. Mexico, Poland, Peru, and St. Lucia). But rules alone do not guarantee compliance (e.g., Albania). Credibility was found to be greater when the government had both a positive track record of strong fiscal measures, and the IMF financial support was in place (e.g. Indonesia, Romania and Latvia). Treatment of risks was sometimes perfunctory and generic, not closely tied to the macro-fiscal and DPO policies in general.
- Sustainability analyses, generally, present credible baseline scenarios that result in the stabilizing or declining debt-to-gross domestic product ratios. In most cases, it is accompanied by a useful scenario analysis showing sensitivity of the baseline debt trajectory in response to growth and interest rate shocks, for example. A question of robustness, however, could be raised in some case studies where there is a lack of complete or clear sensitivity analysis or consideration of various quasi-fiscal risks (e.g., government guarantees,

state-owned enterprises, implicit subsidies) which appear relevant for the operation and may jeopardize sustainability.

- In the reviewed case studies, close collaboration with the IMF were associated with better designed macro frameworks. But there is a question whether the Bank should do more in linking macro-fiscal frameworks to its own prior actions. This is particularly the case regarding tax policy.
- In cases where there is no IMF program or evidence of direct, substantial collaboration with the IMF, and track record is lacking, there appeared to be more weaknesses in certain elements of the macro fiscal frameworks. Greater attention is needed in ensuring the quality of macro-fiscal frameworks in the absence of IMF programs and in the presence of weak track records.
- The length of the time horizon of the operation (standalone or programmatic) affects the capacity to link substantive prior actions to significant, longer-term fiscal and macroeconomic stability results. Standalone operations showed specific weaknesses in the completeness and overall quality of macro-fiscal frameworks.
- The case studies suggest programmatic operations are more suited for certain types of situations (good track record, IMF presence, high quality, and continuity of the dialogue etc.). By contrast, standalone operations appear be better for for reengagement with countries, emergency operations, or sectoral DPOs with shorter time horizons.
- Adding NLTA should be considered with standalone DPOs. This could help build capacity and sustainability and provide the borrower with further analysis on the reforms and their possible fiscal and macro effect in the medium and long run.

III. Insights from the statistical correlation analysis

Insights from a preliminary statistical analysis provide additional evidence of the importance of macro-fiscal frameworks. Some specific findings are consistent with the observations above based on portfolio analysis and case studies. Four findings are highlighted.

First, summary statistics for IEG ratings of DPOs, and indices of the quality of macro-fiscal frameworks and the implementation of macro-fiscal frameworks suggest that (i) IEG ratings of DPOs slightly worsened in the past three years compared to the whole period of analysis, but (ii) the quality of design and implementation of macro-fiscal frameworks improved in recent years.

Second, the elements or indicators of the quality of the macro framework design are positively correlated with the IEG ratings. In particular, other things being equal, the measure of fiscal track record ("backward-looking credibility") and the coverage of quasi-fiscal risks are statistically significantly correlated with the IEG ratings. Of interest is also that programmatic DPOs and DPOs with economic policy thematic focus are positively correlated with IEG ratings, which in some regressions also show statistical significance.

Third, the quality of the macro framework alone (measured along dimensions of consistency, credibility, and sustainability) is not statistically strongly correlated with IEG ratings. However, after controlling for macro implementation quality (measured by the difference between macro-fiscal targets and outturns, with smaller difference indicating better implementation), the macro-fiscal design quality is statistically significantly correlated with IEG ratings.

Fourth, the quality of the macro-fiscal framework matters for implementation:

macro framework design is generally positively correlated with macro implementation.

Some implications for operational guidance and future design of DPOs

In sectoral and standalone DPOs in particular, greater attention should be paid to the macro-fiscal frameworks and the assessment of macroeconomic risks and how specifically risks might affect the sectoral reform agenda. Also, in these operations, there may be a need for more scrutiny during reviews of the completeness of fiscal sustainability and risks that may alter baseline debt scenarios.

Where there is no IMF program or evidence of direct, substantial collaboration with the IMF and track record is lacking, the macro-fiscal framework should receive greater attention in the design and operational review stages (both programmatic and standalone operations) to ensure quality of the macro-fiscal framework and adequate assessment of risks.

The division of labor between the World Bank and IMF should be reviewed with a view towards clarifying their respective roles on tax policy dialogue and analysis. It should provide provide further guidance on the division of labor between the World Bank and the IMF as it relates to DPOs. In particular, it would be useful to review the role of revenue policies, the treatment of which is highly limited in many DPOs.

It could be recognized more explicitly which situations are better suited to programmatic operations (for example: good track record, IMF presence, high quality, and continuity of the dialogue etc.). In other situations standalone operations are better suited (e.g., reengagements, emergency operations, sectoral DPOs with shorter time horizon, special focus DPOs, and DPOs after a new government is in place or after a change in policy direction). Still, there may be cases where special situations (e.g., crisis, new government) warrant flexibility in the choice of the types of DPOs.

Standalone DPOs may benefit from systematically adding accompanying technical assistance (as part of a standalone DPO default "package"). This could help build capacity beyond the horizon of the DPO and help ensure sustainability of DPO impacts.

Parallel, sectoral DPOs, especially in large countries with large DPO programs, may benefit from greater and more systematic scrutiny and articulation of macro-fiscal frameworks and risks. Given the transaction costs of many parallel DPOs, clustering or consolidating such operations should also be considered.

1. Introduction

1.1 The report is part of a series of learning products[4] with limited objectives geared towards synthesizing existing knowledge as well as learning and gaining new insights into the factors that influence design, policy implementation, and performance of Development Policy Loans. It aims to provide fresh insights and build on previous evaluations of Bank investment projects and research and analytical work (e.g., Recent Independent Evaluation Group [IEG] work on Development Policy Operations [DPOs] includes the evaluation of Poverty Reduction Support Credits [PRSCs] [2010], Financial Crisis Evaluations I and II [which examined crisis DPOs], recent Operations Policy and Country Services [OPCS] DPO Retrospectives, and new research presented in this report on aspects of DPOs using a combination of approaches.).

1.2 All DPOs are expected to include an adequate *design* of a macro-fiscal/budgetary framework.

"*The Bank undertakes development policy lending in a country only when it has determined that the country's macroeconomic policy framework is appropriate. The release of each tranche requires the maintenance of an appropriate macroeconomic policy framework. For development policy lending to a subnational entity, the state or region must have an appropriate expenditure program, as well as appropriate fiscal arrangements with the central government. (OP 8.60).*"

1.3 An adequate macro-framework is typically a key condition of Board presentation in one or another form of a specific prior action. Importantly, as noted in OP 8.60, it is also expected that an adequate macro-framework will be *maintained* following Board approval and disbursement of a DPO (in the case of single-tranche DPOs) and this expectation is sometimes formalized in explicit triggers (in the case of programmatic series or multi-tranche DPOs). The last OPCS DPO retrospective noted that the quality of macro frameworks in some DPOs was lacking (see box 2.1). But the quality of macroframeworks in DPOs alone has not been the subject of in-depth study. This report

---

[4] The Learning Products (LP) are IEG informal analytical products, which primarily synthesize existing evaluative material with the aim of gaining new insight into the effectiveness and design of development policies and interventions supported by the Bank from an operational perspective. This learning product series consists of complementary products covering salient areas identified in the 2012 DPO retrospective and in recent discussions of DPO evaluations held jointly between OPCS and IEG. The main evaluative questions relate to the overall effectiveness of DPOs and factors influencing their design and performance. The learning products are organized around several thematic tasks: (i) Adequacy, quality, and risks in DPO macro-fiscal frameworks; (ii) the role of knowledge on Public Expenditure Reviews (PERs) and public resource allocation in DPOs; (iii) results frameworks in DPOs; (iv) environmental and social risks in DPOs; (v) political economy analysis (PEA) and related factors influencing the performance of DPOs; and (vi) the influence of core poverty diagnostics on DPO design and outcomes. The first four tasks are be delivered in FY15 and the remaining ones are planned for FY16. At that time, a synthesis/capstone report will be prepared with main findings and lessons from the learning product series.

aims to contribute to filling that gap with a (i) preliminary analysis of the quality of macro-fiscal frameworks in DPOs, and (ii) a comprehensive database of project level DPOs and related country and economic policy and statistical information, which is a byproduct of the study.

2. What is the adequacy and quality of macro-fiscal frameworks in DPOs?

2.1 For the purpose of this analysis, we adopt the following operational definitions rooted in the Operations Policy and Country Services (OPCS) guidance and regulations and operational practice. Adequacy refers to the statutory requirement that Development Policy Operations (DPOs) must have macro-fiscal framework that meets certain minimum quality requirements in three dimensions: internal consistency, credibility, and debt sustainability (OPCS, September 2013). Quality refers to the degree to which different building blocks of macro-fiscal frameworks can be assessed as having different levels of strength upon careful inspection and analysis. While all DPOs must meet minimum adequacy requirements, the quality of the building blocks and the overall macro frameworks may vary and there can be weaknesses that may not be always obvious in an ex-ante binary, statutory assessment of adequacy. Identifying and improving upon such weaknesses would help contribute to the better design and success of DPOs.

2.2 Consistency, broadly speaking, refers to the degree of correspondence between policy objectives (e.g., macroeconomic stability and sustained growth) and policy instruments (primarily fiscal [tax, expenditure] and monetary-exchange policies) as well as the mix of financing. A consistent program should be complete with sufficient detail, internally coherent, and realistic.

2.3 Where there is an International Monetary Fund (IMF) program in place, the IMF assessment in program reviews provides a key input into the World Bank team's assessment, but the Bank team should make its own assessment (OP 8.60, footnote 4). In cases where there are disagreements over technical or policy issues, Bank and IMF teams are expected to discuss and reach agreement over any differences of opinion and coordinate their policy dialogue. This is facilitated by regular consultations between the Bank and IMF teams and an explicit understanding on the thematic division of labor on macroeconomic and structural issues in each country. However, when there is no IMF program in place, the Bank team's independent assessment is central. In such cases, the IMF may be asked for a "letter of comfort" providing its own view of the macroeconomic policies, but such letters are typically much shorter and less detailed than IMF program reviews (IMF 2003). Hence the Bank team's assessment remains critical.

2.4 Credibility is important for program implementation and success. It reflects the degree to which the government's actions are trusted to be executed in line with pronouncements. Trust in pronouncement is a function of the past track record. A credible program of reforms has a better chance of success because it encourages more desirable and predictable responses of economic agents. Short-term credibility, however, can unwind over a longer period: exchange rate-based stabilizations often result in temporary expansions of activity followed by a likely slump later on, for example (Vegh 2010).[5] This, too, is related to the government's past track record with policy reform. Also, whether or not the government aims at a realistic scale and structure of fiscal adjustment affects credibility. Finally, the IMF program, with its discipline, constraints on macro-fiscal management, and incentives for compliance, can add to credibility. So track record, the realism of the fiscal program, and the presence of the IMF can help strengthen credibility.

2.5 Sustainability—debt sustainability in particular--is a critical element of the adequacy and quality of the macro-fiscal framework. It is assessed through the standard metrics of fiscal and debt sustainability analysis (DSA). An adequate macro-fiscal framework, therefore, must include a medium-term fiscal framework consistent with at least stabilizing or declining debt-to-gross domestic product (GDP) ratios, all backed by a consistent and credible fiscal program. In addition, sustainable fiscal measures are durable or difficult to reverse. For example, a temporary, deep cut in public sector wages or pensions may be an expedient way to reduce the deficit in the short term, but it might prove politically and socially unsustainable. By contrast, a more gradual pace of public sector wage/pension adjustments coupled with difficult but longer-term public sector compensation and pension reform might prove both fiscally and socially more sustainable. So the assessment of the quality of the macro-fiscal framework will need to take into account each of these elements (OP8.60, February 2012, revised July 2014).

Box 2.1. Issues and Recommendations in the 2012 OPCS Three-Year DPO Retrospective

Results: Enhance the focus on sustainable results. For the most part, DPOs have been relatively successful in the achievement of development outcomes and quality appears to be holding despite increased volumes. But there are variations across regions with Europe and Central Asia having the largest number of DPOs rated satisfactorily by the Independent Evaluation Group (IEG) (South Asia and Europe and Central Asia, if the highly satisfactory bar is adopted) and the Middle East and North Africa and Sub-Saharan Africa having the lowest numbers. Programmatic DPOs tend to perform better than standalone DPOs.

[5] For example, if a government announces a blanket deposit insurance scheme in the middle of a currency crisis and deposit outflows without adequate upfront measures and detailed communication to the public, it is not likely to stem the crisis; in fact, it may aggravate it. By contrast, a well thought out and communicated plan with critical upfront actions and timetable of further actions which also convinces the public that the government has the resources, resolve, and intention to deliver on the promised deposit insurance has a much better chance of success.

Economic policy/public sector/poverty/macro DPOs have also performed better than sectoral DPOs. The quality of results frameworks improved compared with the previous retrospective with most DPOs having results indicators with clear baselines and end-of-series targets, and a lower number of indicators showing greater selectivity. Despite this progress, there is room to improve links between policies and results and assess the long-term impacts of DPOs. Monitoring and evaluation (M&E) systems increasingly rely on country systems but only a few DPOs include assessments of strengths and weaknesses of country M&E systems.

Risks and opportunities. The Bank will need to establish standard risk assessment frameworks for DPOs. The Bank has heavily weighted governance and fiduciary risks in the decision to extend DPOs so that a larger share of DPO commitments went to better fiduciary and governance performers. Also, countries with stronger governance and fiduciary systems received a larger share of financing in DPOs compared to investment lending. Where governance and fiduciary risks were high, the Bank addressed these issues in the policy framework and prior actions as measured by the actions devoted to these areas. There is a need to strengthen the quality of macroeconomic assessments. The Bank requires that the macroeconomic framework be consistent with a clear statement of its adequacy.

Reforms. It is recommended to mainstream Policy-Based Guarantees (PBGs) into the operational policy framework governing DPOs (OP 8.60). OPCS will present to the Board a modernized policy on guarantees, including the extension of PBGs to International Development Association (IDA)-only countries with a low risk of debt distress and adequate debt management. Also, the Bank will seek to improve the effectiveness of DPOs under Joint Budget Support (JBS) partnerships.

Source: 2012 OPCS Three-Year DPO Retrospective.

2.6 An adequate macro-framework that passes the basic requirements may be of varying quality. Consistency may be adequate, but fiscal measures may lack sufficient detail-raising questions about the realism of the fiscal program. The track record may be solid, but the scale of fiscal adjustment may raise doubts about credibility. And basic debt sustainability analysis (DSA) may be reported but not in sufficient detail or with sensitivity analysis to account for likely shocks. So the macro-fiscal framework may be adequate, but not "robust" enough in that it fails to account for relevant risks (OPCS 2012 retrospective). A variety of macro, sectoral, and micro risks could derail the implementation of macro frameworks and the broader DPO-supported reforms.[6]

[6] As part of a background analysis, relevant Bank policy regulations and OPCS guidelines (OP/BP 8.60) have been reviewed. Where possible, insights from in-depth Project Performance Assessment Reports (PPAR) were used (e.g., Tanzania, Vietnam, and Uganda). A number of OPCS documents (e.g., OP/BP 8.60; OPCS 2013; OPCS DPO Academy materials; DPO retrospective 2012, OPCS Guidance Note on macroeframeworks, 2013), IEG (e.g., Evaluation of PRSCs 2010; Crisis DPOs, Independent Evaluation Office (IEO) of the IMF, and research studies (e.g., Moll, Geli, and Saavedra 2015; Geli, Kraay, and Nabokht 2014; Smets and Knack 2014; Dobronogov, Gelb and Saldanha 2014; Denizer, Kaufmann and Kraay 2011; Dreher et al. 2010; Bogetic et al. 2010; Ossowsky 2014) relevant for this learning product have been consulted. These analyses provide useful background and touch on various

positive or normative aspects of the effectiveness of World Bank lending more generally, and key aspects of macroeconomic policy frameworks or DPO lending in particular.

3. Questions and analytical approaches

3.1 The three overarching questions in this paper are:

- What has been the quality of macro-fiscal frameworks and its building blocks?
- What factors may be related to the quality of macro-fiscal frameworks and the success of DPOs?
- What good practices can be identified that may provide lessons for the future design and implementation of DPO macro-fiscal frameworks?

3.2 We approach these questions using an eclectic analytical approach combining:

- **elements of a portfolio review of DPOs** with particular focus on the aforementioned dimensions of macro-fiscal frameworks and their links to macro-fiscal policy reforms supported under DPOs,
- **a comparative review based on 15 case studies** providing deeper narratives of the quality and factors that influence the design and implementation of macro frameworks and their contribution to DPO performance, and
- **preliminary statistical analysis** to understand the macroeconomic framework-related correlates of success of DPOs, as well as the correlates of the quality of macro-fiscal framework itself.

3.3 Insights from these analyses were triangulated and conclusions were drawn. This includes some thoughts toward improving the design and implementation of macro-fiscal frameworks in future DPOs as well as questions for further research.

3.4 The target pool of DPOs and the periodicity under study is the following: first, all DPOs that were closed and had an Implementation Completion Report (ICR) review during FY05-13—a total of 390 DPOs—were included in the database of basic characteristics of DPOs. This pool of DPOs and related data are used in the preliminary portfolio review and statistical analysis. Second, closer attention was paid to the 46 DPOs that closed and had ICR reviews completed in FY11-13. About a third of these recent DPOs (15 in total) were selected for in-depth, desk-based country case studies to uncover deeper country-level and project-level narratives and elements of good practice (and weakness) in the quality of macro frameworks. Case studies were non-randomly selected from the pool of the most recent, 46 DPOs that were closed and evaluated in FY2011-13 to reflect diversity of DPOs in terms of programmatic and standalone DPOs, multi-sector and sector DPOs, and DPOs reflecting all world regions. A comprehensive Independent Evaluation Group DPO database was developed for this purpose and it informed this learning product and other DPO learning products in the series.

4. DPOs and macro-fiscal frameworks: A basic portfolio view

4.1 This section provides a basic portfolio review of Development Policy Operations (DPOs) analyzed in this learning product. It focuses on the following main elements:

- A general review of trends in volumes and numbers of DPOs, as well as their country/income groupings (International Bank for Reconstruction and Development [IBRD], International Development Association [IDA], Blend) and regional focus.
- A review of the policy and thematic focus of DPOs as reflected in their prior actions.
- A review of basic metrics of the integration of macro-fiscal topics and prior actions in DPOs.

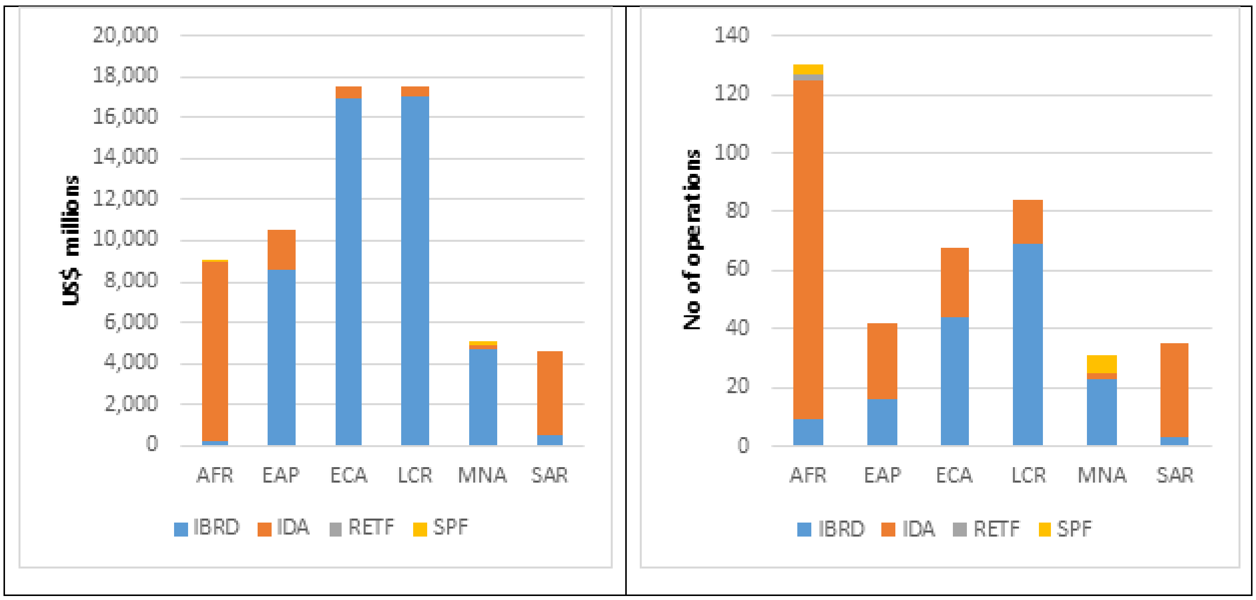
Trends in the volumes, numbers, IBRD/IDA, and geography of DPOs

4.2 **Method**. For identification purposes, the detailed Bank project theme report was downloaded and customized from Business Warehouse in August 2014 (2c.2.1 lines). Only DPOs were retained from the database. Independent Evaluation Group (IEG) Implementation Completion Report (ICR) Review outcome ratings were merged and matched with the DPO database. Observations for DPOs without an IEG validation were dropped. Based on this selection process, 390 DPOs were identified for the period of 2004-2013. Next, a database of prior actions was received from Operations Policy and Country Services (OPCS), matching 387 of the 390 DPOs identified through Business Warehouse. Thematic codes were further reviewed manually by IEG and broken down into a few subtopics to improve the granularity of the analysis of macroeconomic management issues in DPOs. Two main limitations apply to the underlying data. First, administrative data for a given project are typically recorded at an early stage of preparation, and the record is unlikely to be changed even if there may be significant changes later on. While the thematic focus of projects is typically expected to remain robust from preparation and concept to board approval, this may have led to some inaccuracies when the project focus was significantly changed during preparation. Second, the number of observations for recent years is significantly lower, as there has been a delay in the review of closed projects. The latter is also related to technical problems in transferring ICRs to IEG for review.

4.3 **Volumes, numbers, and regional distribution**. Two Regions, Europe and Central Asia and Latin America and the Caribbean have received the most DPO lending commitments in terms of volume (almost $18 billion each). Most were IBRD countries. This is followed by East Asia and the Pacific, which has proportionately more IDA countries. Next is the Africa Region, with mostly IDA countries.[7]

4.4 In terms of number of DPOs, the Africa Region has the more DPO operations than any other Region, partly reflecting the large number of countries. The average size of the DPO (in terms of US$) in the Africa Region, however, is significantly lower than that of other Regions, reflecting the comparatively smaller size of African economies and their absorptive capacity (figure 4.1).

Figure 4.1 Commitments and numbers of DPOs by lending instrument and region (closed DPOs, 2004-2013)

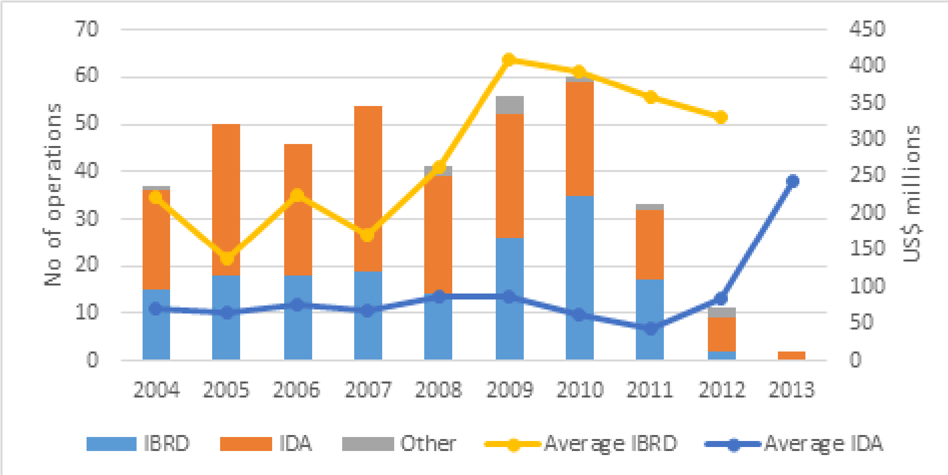


Source: IEG DPO database.

4.5 The average lending amount for IDA countries has remained largely constant over time, but the lending amount increased dramatically for IBRD countries in the late 2000s. This increase was a result of the global financial crisis when DPOs were used as a countercyclical instrument in IBRD countries. The reversal of these trends in 2012 and 2013 may simply reflect the low number of observations (low number of operations closed and reviewed) and should be viewed with caution (figure 4.2).

[7] There were two recipient executed trust funds (RETFs). One in Rwanda (2010) and one Burkina Faso (2012), and eight DPOs (e.g. West Bank and Gaza and Burundi) using the state and peace building fund (SPF). These constitute a minor part of the portfolio, both in terms of numbers and volume.

Figure 4.2 Trends in IBRD and IDA DPOs and average commitments (closed DPOs 2004-2013)



Source: IEG DPO database.

4.6 The top recipients of DPO financing by volume are IBRD countries. Most notably Mexico and Indonesia each received more than US$ 6 billion, followed by Poland and Turkey, which received close to US$ 5 billion each. The countries that received the largest number of DPOs were Pakistan (16 operations) and Colombia (14 operations) (Table 4.1).

Table 4.1 Top ten countries by commitments and number of operations (closed operations 2004-2013)

#	Top ten countries by commitment	Total commitment (US$ millions)	Top ten countries by no of operations	No of operations
1	Mexico	6,916	Pakistan	16
2	Indonesia	6,250	Colombia	14
3	Poland	4,996	Mexico	13
4	Turkey	4,900	Morocco	13
5	Colombia	3,650	Vietnam	11
6	Pakistan	2,760	Indonesia	11
7	Vietnam	2,175	Bangladesh	10
8	Ukraine	1,701	Ghana	9
9	Morocco	1,638	Peru	9
10	Romania	1,514	Mozambique / Tanzania	8

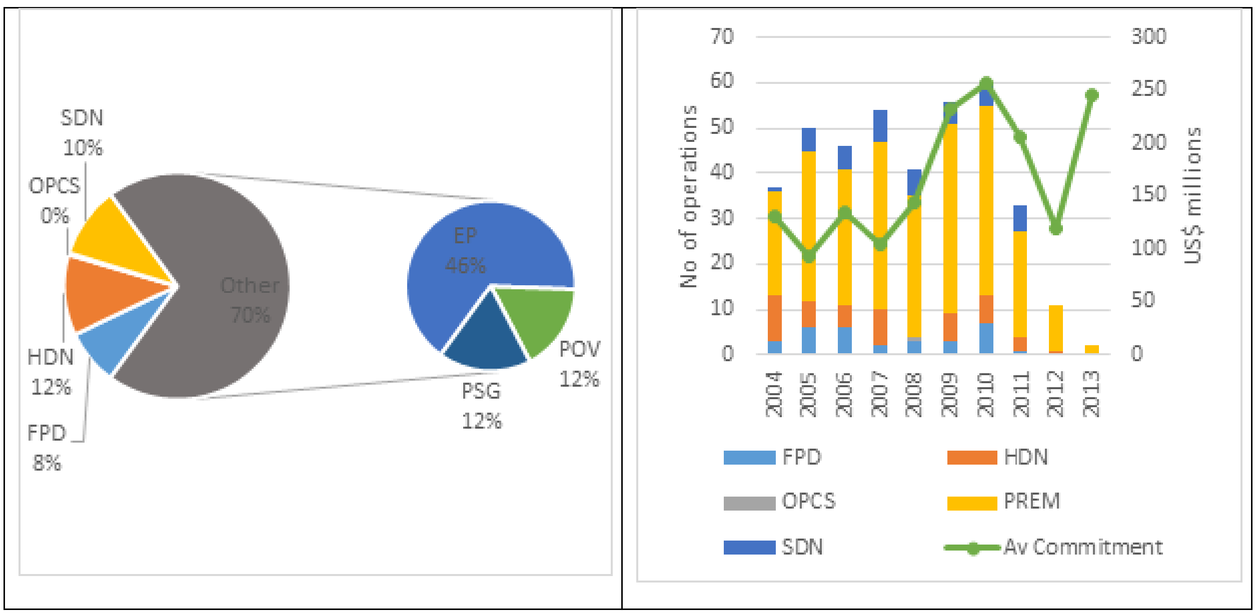
Source: IEG DPO Database.

Trends in the types of DPOs

4.7 The majority of DPOs (70 percent) were managed by the Poverty Reduction and Economic Management Network, followed by Human Development (HD) (12 percent) and Sustainable Development Network (10 percent). Within PREM, most DPOs were managed by economic policy. The Poverty (POV) and Public Sector (PS) practices each managed 12 percent of all DPOs.

4.8 There does not appear to be substantial variation over time, regarding the networks managing DPOs (Figure 4.3). This may reflect, on the demand side, the thematic focus of reforms supported by DPOs, typically concentrated in economic policy and governance issues and, on the supply side, the Bank's institutional tendency towards a relatively constant division of labor among networks. The notable reduction of PREM-led DPOs from 2011 onward, however, reflects the low number of closed and reviewed operations rather than a trend in the reduction of approvals.

Figure 4.3. Distribution and trend of DPOs managed by Networks and Sector Boards (closed operations 2004-2013)



Source: IEG DPO Database.

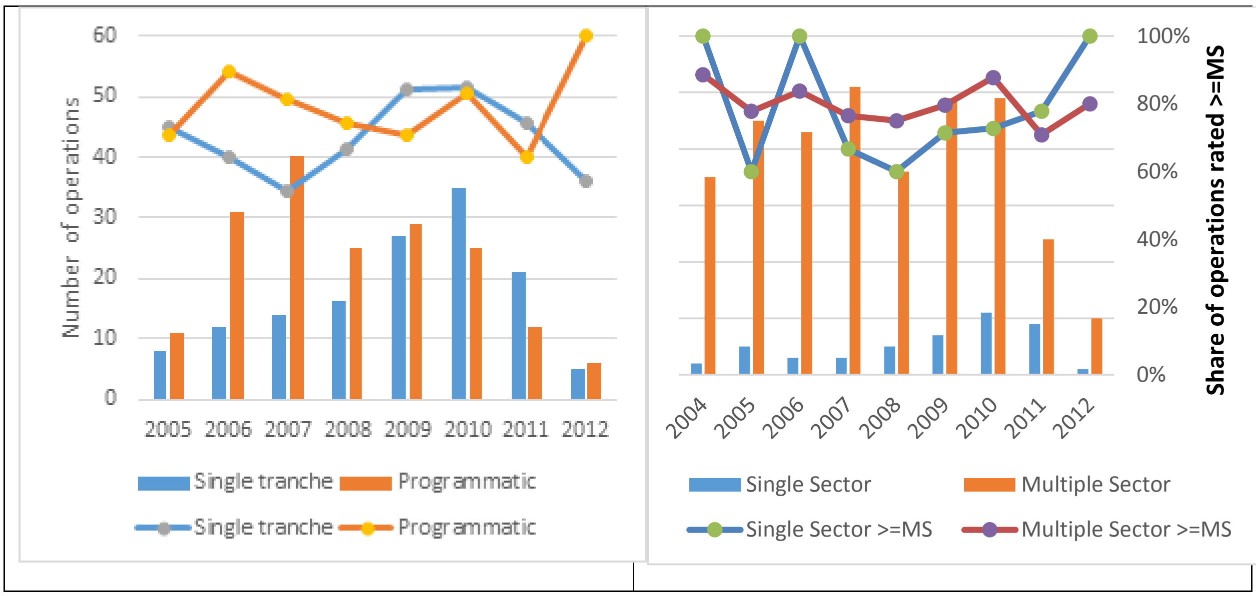
4.9 Of the 390 DPOs, 47 percent are programmatic and 42 percent standalone. In the early years after the OP8.60, there were substantially more programmatic DPOs than single tranche. After 2008, there has been a steady increase in standalone operations. (12 percent were prior to OP8.60).

4.10 Programmatic DPOs appear to have performed better in terms of IEG ratings than standalone operations between 2005 and 2008. Thereafter, performance was relatively equal. About 80 percent of all DPOs were rated moderately satisfactory or above.

4.11 Most DPOs were multi-sector operations (they could be programmatic or standalone), but single-sector operations seem to be increasing. The metric used to identify the sectoral focus of DPOs in the IEG DPO database used here is the following: a DPO is coded for a single sector and 75 percent or more were identified as such.

4.12 Multi-sector DPOs have been performing better than single sector DPOs over the years. The exception years are 2004 and 2006, which appear to be outliers. The trends in these DPOs is displayed in figure 4.4.

Figure 4.4 Trends and ratings by types of DPOs (closed operations 2004-2013)

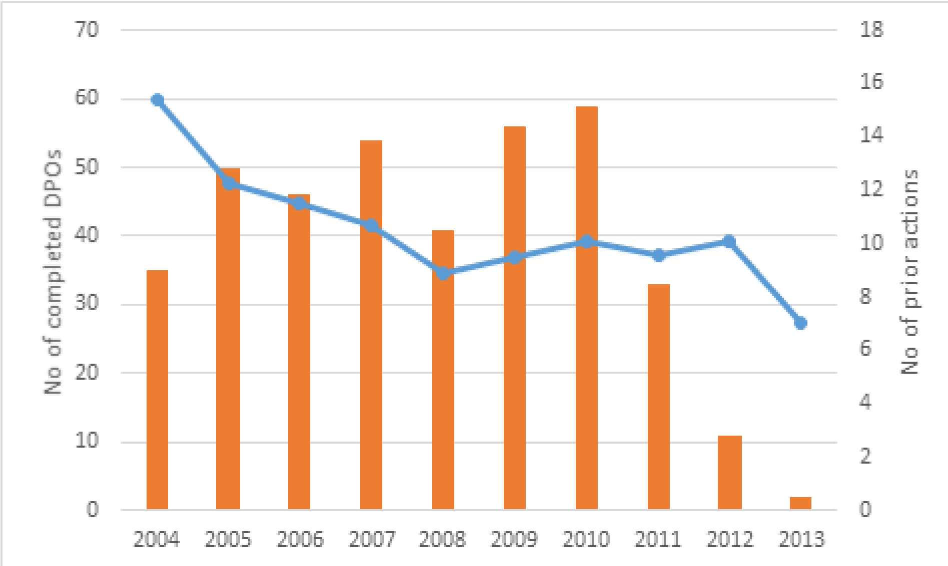


Source: IEG DPO database.

Trends in the policy focus of DPOs: prior actions

4.13 The average number of prior actions fell from 16 to nine between 2004 and 2008. This is in line with the OPCS guidance on simplification, reducing the number and increasing concreteness, relevance, and quality of prior actions. After 2008, the average number has stabilized between eight and 10 (Figure 4.5).

Figure 4.5 The number of prior actions per DPO (closed operations 2004-2013)



Source: OPCS prior actions database.

4.14 The majority of all prior actions are in the domain of public sector governance (43.5 percent), followed by financial and private sector development (18.3 percent), and human development (10.5 percent). OPCS maintains a list of prior actions, and codes these systematically against protocol. These themes are clustered and summarized in table 4.2.

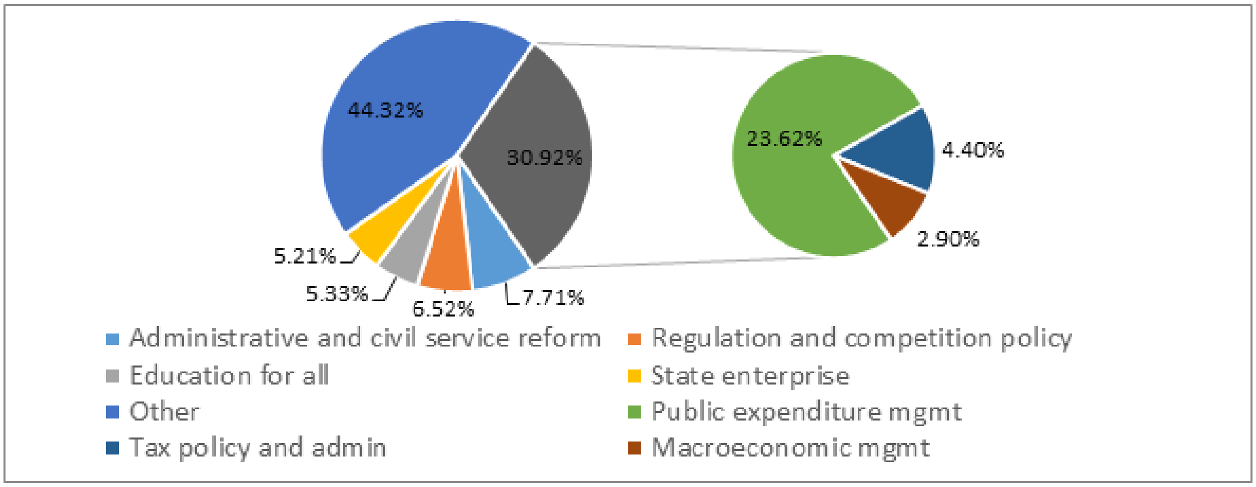
Table 4.2 Distribution of all prior action themes by clusters (closed operations 2004-2013)

Theme code clusters	Share
Public Sector Governance	43.5%
Financial and Private Sector Development	18.3%
Human Development	10.5%
Social Protection and Risk Management	6.7%
Environment and Natural Resources Management	5.0%
Economic Management	4.7%
Trade and Integration	3.1%
Rule of Law	2.9%
Rural Development	2.7%
Urban Development	1.3%
Social Development, Gender, and Inclusion	1.2%
MACRO-FISCAL THEMES	
Public expenditure, financial management, and procurement	23.6%
Tax policy and administration	4.4%
Macroeconomic management	2.9%

Source: OPCS Prior Actions database.

4.15 Macro-fiscal and public financial management (PFM)-related prior actions in DPOs make up about 30 percent of all prior actions. Within this macro-fiscal and PFM, public expenditure, financial management, and procurement is by far the largest category (23.6 percent), followed by tax policy and administration (4.4 percent), and macroeconomic management (2.9 percent). Other prior actions include administrative and civil service reform (7.7 percent), regulation and competition policy (7.0 percent), and education for all (5.3 percent) (figure 4.6).

Figure 4.6.Top thematic areas of prior actions (left chart), and macro-fiscal related themes (right chart) (closed operations 2004-2013)



Source: IEG DPO database based on the OPCS prior actions database.

4.16 Given the majority of PREM-managed, multi-sector operations, the comparatively small number of macro-fiscal actions and tax policy and administration (2.9 percent and 4.4 percent of all prior actions) is unusual. While macro and tax-related prior actions are few as a share of the total, they are better distributed by projects. In macro, 90 out of 387 projects contain at least one prior action, which is equivalent to 23 percent. Similarly 110 DPOs contain at least one prior action in tax policy and administration, an equivalent of 28 percent of all operations. Almost half of all operations (45 percent) have either macro or tax policy as one prior action. About 70 percent of all operations containing macro actions were managed by PREM. Operations that do not have such prior actions may reflect the fact that the Bank left many of the critical macro and tax policy measures to the International Monetary Fund (IMF) programs. Only in four instances was cross conditionality with the IMF reported in the entire set of macro prior actions.

4.17 The implementation of an adequate macroeconomic framework, or a close derivative (e.g., PFM), was most frequently used as a macro-economic prior action. Of all prior actions with macro-fiscal and PFM content, some 69 percent of prior actions targeted broad maintenance of sound macroeconomic policy, followed by 27 percent of macro-fiscal actions specifically targeting the budget (e.g., submission of the budget or its amendment) and 19 percent targeted concrete expenditure policy measures. Tax policy is covered in large part under tax policy and administration (see figure 4.6), and is thus not well represented in measures classified as macro-fiscal. However, a preliminary inspection of the tax policy and administration of prior actions suggest relative paucity of tax policy and dominance of administration measures. More than half of public expenditure policy actions targeted the medium term expenditure framework (MTEF) explicitly, reflecting the focus of DPOs on medium-term expenditure allocation and management issues.

4.18 In terms of the nature of macro-fiscal prior actions, most relate to the policy implementation (75 percent). Prior actions that are documentary (e.g. the Ministry of Finance has issued a report) were 21 percent. Legislative reform (e.g., preparing, submitting, or passing of a bill) account for 10 percent of all macro-fiscal actions (table 4.3).

4.19 Given that many DPOs target substantial fiscal adjustments, the relative absence of explicit prior actions on tax policy is puzzling and requires further study. It may be that detailed tax policy dialogue and conditionality is left to the IMF, even though tax policy is a shared thematic area of responsibility between the Bank and the IMF. It may also be that some of the tax policy measures are part of the package of other prior actions targeting an "adequate budgetary framework" or a "sound macroeconomic policy framework." The scope, structure, and the role of tax policy in DPOs and in relation to the IMF would be a worthwhile extension of this analysis.

Table 4.3 Target policy and type of macro-fiscal prior actions (closed operations 2004-2013)[8]

	No of actions	Total actions in macro fiscal area	Share of actions
PA TARGET			
Macroeconomic policy	85	123	69%
Budget	33	123	27%
Tax policy	1	123	1%
Expenditure policy / PFM	23	123	19%
MTEF	14	123	11%
Other / quasi fiscal(1)	9	123	7%
PA TYPE			
Implementation	92	123	75%
Legislative	12	123	10%
Documentary	26	123	21%

Source: IEG review of OPCS prior action data.

[8] Shares exceed 100 percent as prior actions could be tagged with multiple codes.

Consistency in the macro-fiscal framework

4.20 **Method**. The program documents for 174 of the 390 operations (45 percent) were reviewed and assessed for consistency, a key element of quality of the macro-fiscal framework. This sample, chosen for practical reasons and the time constraint, significantly exceeds the usually required size of the random sample. The focus was on the macroeconomic framework and debt sustainability parts of the program documents. The team developed a 5-point scale (1 for low, 5 for high) to code independently for the degree of consistency of the macro-fiscal framework. The program documents of these operations were screened for the completeness of the basic macro-fiscal framework, detail and coherence between macro-fiscal objectives and fiscal measures, and coverage of other elements of the macro-framework including the external sector, monetary exchange, and, especially debt sustainability. A score of 5 was allocated to operations with a detailed, in-depth, macroeconomic outlook, with clear and credible links between macro-fiscal objectives and detailed measures, and debt sustainability assessment across all of the above dimensions, and a score of 1 was given to operations where completeness was clearly lacking in a major or several dimensions of the macro-fiscal framework. The mid-point of 3 was given to operations showing solid consistency in terms of basic requirements.

4.21 **Results**. Using this approach, the results show that the consistency of the macro-fiscal framework was solid, rated 3 or better, for the majority of operations. This is reassuring as the majority of operations pass this independent test of consistency of the macro-fiscal framework. It suggests that the operational framework and review processes work well in screening for the quality of the macro-fiscal framework in most operations. The average score was 3.4. While the mid-point (3) was most frequent (28 percent), this was closely followed by DPOs with substantial (25percent) or high (21 percent) levels of consistency. About 22 percent of operations scored modestly (rating 2), typically showing weakness in one important area of the macro-fiscal framework. This analysis is agnostic about whether these DPOs have adequate macro-fiscal frameworks in the statutory sense. There may be other reasons (including, for example, paucity of data, the emergency nature of the operation, etc.) that may justify these weaknesses. But this independent review focusing on the overall and relative quality identifies certain weaknesses, which could be addressed in future operations. Such weaknesses, for example, include the absence of sufficiently detailed measures in support of the macro-fiscal objectives; weak revenue performance not matched by tax policy or administration measures; or debt sustainability analysis that is insufficiently detailed. Only 4 percent of operations show negligible consistency (rating 1). By Regions, East Asia and the Pacific stands out with an exceptional number of operations with a high consistency rating (38 percent). However, there are DPOs across the spectrum of the consistency index in all the Regions. Table 4.4 provides a detailed breakdown by Regions and scores.

Table 4.4 Consistency score of 174 DPOs (2004-2013)

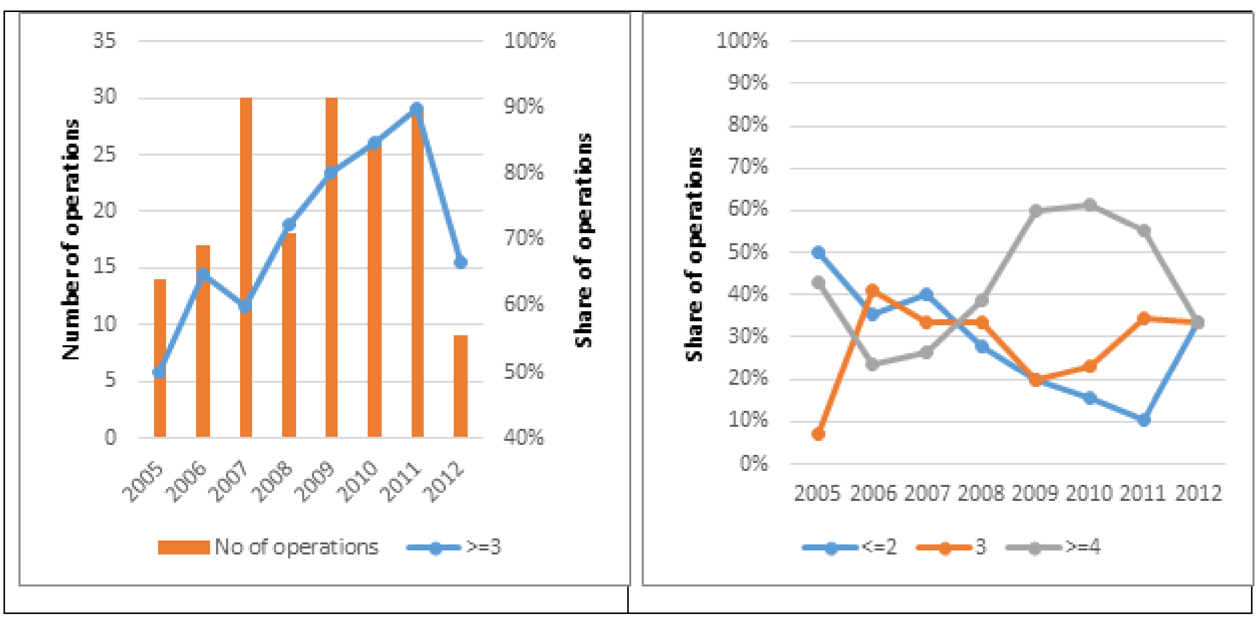
Region	1 (low)	2	3	4	5 (high)	Total
AFR	2,(3.4%)	14,(23.7%)	15,(25.4%)	20,(33.9%)	8,(13.6%)	59,(100%)
EAP	0,(0.0%)	5,(20.8%)	7,(29.2%)	3,(12.5%)	9,(37.5%)	24,(100%)
ECA	0,(0.0%)	6,(20.0%)	11,(36.7%)	8,(26.7%)	5,(16.7%)	30,(100%)
LCR	2,(7.1%)	7,(25.0%)	6,(21.4%)	4,(14.3%)	9,(32.1%)	28,(100%)
MNA	0,(0.0%)	4,(22.2%)	3,(16.7%)	7,(38.9%)	4,(22.2%)	18,(100%)
SAR	3,(20.0%)	3,(20.0%)	7,(46.7%)	1,(6.7%)	1,(6.7%)	15,(100%)
Total	**7,(4.0%)**	**39,(22.4%)**	**49,(28.2%)**	**43,(24.7%)**	**36,(20.7%)**	**174,(100.0%)**

Source: IEG analysis of DPO Program Documents.

4.22 This measure of consistency shows improvement over time. There has been a steady and significant increase in the share of operations scoring 3 or above over the years using this metric. Only in 2012 there appears to be somewhat of a decline, but this may reflect the small sample size and this would need to be checked when more recent operations are included in the updated analysis.

4.23 Beneath this aggregate measure, there is considerable variation over time. Between 2005 and 2007, significantly more operations had lower consistency. This trend was reversed in 2008. Between 2008 and 2011, more operations scored exceptional consistency than those below the mid-point score. In 2012, the shares converged. The share of operations with a mid-point (3) consistency score remained relatively constant in comparison (see figure 4.7).

Figure 4.7.Trends in macro-fiscal consistency score



Source: IEG analysis of DPO Program Documents.

4.24 With these preliminary findings and basic portfolio and consistency metrics, we turn to an in-depth review of 15 case studies, accounting for about a third of recent DPOs from FY11-13 which also have a completed ICR review.

5. Insights from Desk Reviews of 15 Recent DPO Case Studies

5.1 This section draws some initial conclusion on the adequacy of the macroeconomic and fiscal framework from a set of 15 Development Policy Operations (DPOs)—or series—in 14 countries, which were closed and had an Implementation Completion Report (ICR) review in FY11-13. This represents a third of all DPOs in this period having been closed and reviewed. Eight of the 15 operations reviewed were part of a DPO series, while the rest were standalone operations. The DPOs reviewed are presented in table 5.1.

Table 5.1 Countries and DPOs reviewed for case studies

Country and Operations	Period covered
St. Lucia: **Economic and Social Development Policy Credit**	2011-2012
Cote d'Ivoire: **Post Reconstruction and Recovery Grant**	2012-2013
Indonesia: **Infrastructure DPL series**	2007-2010
Peru: **Fiscal Management and Competitiveness DPL series**	2006-2011
Tunisia: **Governance and Opportunity DPL**	2011-2012
Latvia: **Social Safety Net Reform Program**	2010-2011
Poland: **Energy efficiency DPL**	2011-2012
Romania: **DPL series**	2009-2011
Benin: **PRSC series**	2007-2010
Brazil **(Rio Grande do Sul): Fiscal Sustainability for Growth Program DPL**	2009-2011
Albania: **Social Sector Reform Program DPL**	2011-2012
Vietnam: **PRSC series, PIR and P135-2**	2007-2012
Mexico: **Green Growth DPL**	2009-2010
Uganda: **PRSC series 8-9**	2008-2011

Source: IEG team review.

5.2 The case studies were selected to reflect a variety of different types of DPOs. The sample includes all World Bank Regions; International Development Association (IDA) and International Bank for Reconstruction and Development (IBRD) countries; programmatic and standalone DPOs; and a variety of multisector and sectoral DPOs, including, for example, "green growth," infrastructure, and social safety net DPOs.

5.3 The case studies are desk review exercises. They evaluate the following questions organized around the four thematic building blocks corresponding to specific DPO learning products: (i) macro-fiscal framework, (ii) the role of knowledge on public expenditure and resource allocation, (iii) knowledge on poverty, and (iv) environmental and social risks.

5.4 The documentary evidence reviewed comes primarily from the key project, project validation, and project evaluation documents and related, strategic and International Monetary Fund (IMF) documents: the program document (PD), the ICR, ICR reviews (ICRRs), Public Expenditure Reviews (PERs), and Poverty Assessments (PA), if they were available. As secondary sources of evidence, the country case studies (CCS) reviewed the IMF Article IV consultation, Country Economic Memorandums (CEMs), and any other relevant strategic analytical work, such as Country Partnership Strategies (CPSs) or Poverty and Social Impact Analyses (PSIAs).

5.5 This analysis below synthesizes insights from these case studies focused on the consistency, credibility, and sustainability of the macro-fiscal framework.

Main messages

5.6 Overall, the macro-fiscal frameworks presented in DPO desk reviews are broadly adequate. Basic macro frameworks in most cases appear consistent, credible, and sustainable based on the facts presented in the program document as well as related evidence reviewed (e.g., IMF reports, detailed debt sustainability analysis).

5.7 In most cases, the macro-fiscal framework is highly aligned with the IMF and related strategic lending and analytic and advisory activities (AAA). This includes IMF program documents or Article IV consultations, as well nonlending technical assistance (NLTA), economic and sector work (ESW), and other AAA relevant work (a good example is the Peru DPO). From the 15 DPOs reviewed, this overall conclusion holds for the majority of the cases (13 cases). In these cases, the macro framework presented in the program document is consistent with the data presented, as well as with the IMF program documents or Article IV evaluation of the economy (e.g., Peru, St. Lucia, Indonesia, Latvia, Poland, and Romania), reflecting collaboration of Bank and IMF teams (IMF and World Bank, 2011). In many cases, macroeconomic targets are broadly realistic and consistent with the main measures designed to achieve them (good examples are Peru and Romania).

5.8 At the same time, some weaknesses in the quality of macro frameworks are identified in a few operations. Three broad types of weaknesses are highlighted: (i) the ambitiousness of macro-fiscal frameworks in some standalone operations and links between objectives and fiscal measures, (ii) the credibility of the framework in view of the track record, political economy factors, or institutional fiscal rules, and (iii) robustness of the debt sustainability analysis.

- For example, in some cases, the targets seem to be disjointed in terms of the level of ambition in objectives and timing of what the operation is supporting.

In standalone operations with multiple objectives, for example, overly ambitious macro targets are hard to achieve within the time envisaged (e.g., St. Lucia). In other cases, fiscal measures in support of objectives are not sufficiently articulated or left implicit.

- In terms of the credibility of the macro and fiscal frameworks presented in the program documents, a strong track record of fiscal prudence and an existing robust institutional framework, such as fiscal rules, tend to coincide with a better ability to advance reforms while keeping the finances in good shape (e.g. Mexico, Poland, Peru, and St. Lucia). But rules alone do not guarantee compliance (e.g., Albania). Credibility is generally greater when the government has both a positive track record of strong fiscal measures, and the IMF financial support is in place (e.g. Indonesia, Romania, and Latvia). Risks are sometimes generic and not sufficiently tailored to the policy content of the DPO or macro-fiscal situations described therein.
- Sustainability analyses generally present credible baseline scenarios that result in the stabilizing or declining of debt-to-gross domestic product (GDP) ratios. In most cases, it is accompanied by a useful scenario analysis showing sensitivity of the baseline debt trajectory in response to growth and interest rate shocks, for example. A question of robustness, however, could be raised in some cases where there is a lack of complete or clear sensitivity analysis or consideration of various quasi-fiscal risks (e.g., government guarantees, state-owned enterprises, implicit subsidies), which may jeopardize sustainability.

Weak and strong elements of macro-fiscal frameworks

5.9 Beyond these general conclusions, there are specific examples where the elements of quality in macro-fiscal frameworks could be improved.

- For example, the mere presence of fiscal rules does not guarantee compliance or the implementation of prudent fiscal policy. When fiscal rules are in place but the borrower government has not shown a strong track record of following its own fiscal rule legislation, the macro framework could still lack credibility. In these cases, prior actions and triggers could reflect the need to ensure implementation of the existing legislation, ensure corrective action in the case of non-compliance with the rules, or strengthen the relevant institutions (e.g. Albania, Rio Grande do Sul in Brazil and Romania). The latter could be done via complementary, targeted technical assistance.
- Additionally, governments' conflicting objectives may create difficult tradeoffs in terms of fiscal management during implementation and can be a cause of weaknesses in macro-fiscal frameworks. "Austerity" versus protection of the

poor is a case in point. For example, improving public sector governance and management based on cutting waste and increasing taxes—as key elements in fiscal consolidations—may conflict with the need to improve the effectiveness of social safety nets, protect the poor, and combat poverty, which could call for higher, pro-poor government expenditures. If such policy trade-offs are not explicit with measures aiming to strike an appropriate balance, policy objectives and measures might be in conflict. Furthermore, the reasons why the government may find it difficult to strike an appropriate balance here tends to be highly political or related to political economy. A party in power may rely on votes of the public sector whose salaries and pensions have increased and the government may find it politically difficult or impossible to control public sector salaries and shift the composition of public expenditures towards the poor. So the political economy behind these objectives play an important role and should be further studied. Also, some important reforms may not be addressed due to the alignment of reforms with elections and this could be playing a key role in the design and quality of the reform agenda. These issues of the links between the DPOs, their performance, and political economy might require further study, but are outside the scope of this analysis.

- Another example of problems in the design of macro frameworks is that some of the risks and vulnerabilities are not considered adequately. Improving the treatment of risk in macro frameworks has been identified as an issue in the 2012 DPO Retrospective. So this is an area for continued attention. Such risks may be external—for example, a high dependency on external accounts, a significant reliance on tourism and foreign direct investment, and exposure to natural disasters. This vulnerabilities affect the probability of success in implementing fiscal measures and the overall sustainability of the macro framework. Risks may be policy related but not adequately taken into account (e.g., measures to freeze or reduce public sector employment without taking account of potential increases in public sector wages or measures to reduce wages but nothing on employment). Finally, there may be internal Bank pressures to prepare the loan rapidly, perhaps due to the financial crisis, which results in elements of the macro-framework not being fully articulated (e.g., Cote d’Ivoire, Tunisia, and St. Lucia); this may especially be the case in standalone or single-tranche operations.
- In the reviewed case studies, close collaboration with IMF seems to be associated with the better design of macro frameworks. But there is a question whether the Bank should do more in linking macro-fiscal frameworks to its own prior actions. The analysis in the program documents and IMF information, program, and Article IV consultations, are typically consistent in terms of priorities and challenges. In many cases, where an IMF program is present,

however, several macro prior actions or structural reforms affecting fiscal sustainability are left for IMF to address, while the Bank program is silent (e.g., Benin, Latvia, and Romania). In other cases, prior actions are linked with the medium-term expenditure policy agenda through formulation or implementation of medium-term expenditure frameworks.

- In cases where there is no IMF program or evidence of direct, substantial collaboration with IMF, weaknesses appear in certain elements of the macro-fiscal frameworks. For example, in Tunisia, the Bank's operation supported the expansion of several programs without a proper fiscal sustainability analysis and regard for economic vulnerabilities that the country might face in the near future. Similarly, in Brazil's Rio Grande do Sul, the operation showed some inconsistencies between the measures supported and the apparently over-optimistic fiscal projection. The macro framework and track record in this case appear not to be strong enough or credible to validate the projections.

5.10 In what follows, we divide the case studies into three groups for more granular views of strengths and weaknesses of the DPOs reviewed: DPOs of countries in the middle of a fiscal crisis (mainly fiscal DPOs), DPOs combining multiple fiscal and structural objectives, and, finally, DPOs with a mostly sectoral focus.

Table 5.2.Case studies of DPOs in the midst of fiscal/macro unsustainable imbalances

I.	DPO's in the middle of fiscal/macro unsustainable imbalances			IMF Program Present
DPO		Macro/Fiscal Measures supported by Prior Actions		
	DPL rationale and clarity of macro diagnostics	Consistency	Institutional depth	
Brazil RGS	Rio Grande do Sul has had significant problems with fiscal sustainability; it is one of the two states that did not comply with the Fiscal Responsibility Law. The reform agenda covered revenue, pensions and excessive wage bill. The program document is candid and comprehensive.	The operation presents some inconsistencies between the very optimistic projections of fiscal improvements and the measures supported. The measures are too vague and not strong enough to validate the projections.	The measures supported have little institutional depth and all focus on quantity measures without listing or proposing deeper institutional changes. Thus, the prior actions are of poor quality, both on the revenue and the expenditure side, to assure the progress being envisaged.	No
Romania	Romania has showed previous pro-cyclical public and private spending creating a large and increasing debt liability in both sectors of the economy. The global crisis resulted in a credit crunch, large macro imbalances, and a rollover crisis. The PAD presents a very clear case for the financing needs. PAD features strong macro fiscal analysis.	Many areas flagged as a source of pro-cyclical spending are not addressed. Financial sector regulation, civil service reform, and SOEs arrears. Several of these issues were left to the IMF program.	Little institutional depth in Prior Actions selected. The main PA in the fiscal areas include developing a draft MTEF to be eventually approved by parliament and expenditure limits incorporated into the draft. The FRL in place is not being respected. On the wage bill side, ordinances on wage adjustments have been issued, but there is no mention of civil service reform to reduce long-term fiscal presure.	Yes
Tunisia	This emergency operation aimed to support the interim government of Tunisia, in the aftermath of the Arab Spring. Tunisia ha also been hit by the Libyan crisis, leading to lower FDI, and remittances as well as a contraction in exports, creating several budgetary gaps. The IMF is not present. No explanation provided why the Fund was not operating in Tunisia at the time of the operation. Collaboration and risk sharing could have been emphasized.	The program supported the expansion of several social programs without the proper sustainability analysis. The documents reviewed suggest that the Bank was over optimistic regarding Tunisia's reforms, disregarding challenges of implementation.	The PAs focus on governance, access to information, procurement issues and employment programs. There are no PA on the fiscal side, nor messages regarding future measures to rationalize government expenditures.	No
Cote d'Ivoire	The PAD presented a comprehensive discussion of the macro framework. It also provided a thorough debt sustainability analysis. Collaboration with the Fund seems crucial in this operation since the country is a post-conflict state with weak institutions. It has no fiscal rules in place.	All fiscal and structural reforms have been covered under the Fund's program, including wage bill reform. There is a question of value added of Bank supported policy reforms.	The only macro/fiscal PAs associated with this operation is the sectoral MTEF for health and education and the budget execution law. These PAs do not have enough institutional depth and do not represent sufficient additionality to advance policy reforms.	Yes

Source: IEG team review.

5.11 The quality of the links between macro-fiscal design and the reform agenda and prior actions of these operations are mixed. The lack of structural depth in the measures supported in the Brazil Rio Grande do Sul operation resulted in a disconnect between the severity of the problem and the actions taken. Several of the actions taken were administrative initiatives of narrow technical character and easily reversible, compromising credibility and sustainability of the reforms. The passage of key pension reform legislation did not result in a much needed legislation to contain public expenditures. According to the ICRR, progress in social security, civil administration, and results-based management were at best mixed. The ICRR rating of this operation was moderately satisfactory.

5.12 In the Romania DPO, design of the macro framework was particularly strong and comprehensive. The program document identified several of the sources of fiscal imbalance, however, several of the issues raised were left in the charge of the IMF program. Prior actions selected in the operation lacked institutional depth, and were easily reversible, thus hurting the credibility of the reforms in terms of the macro and fiscal sustainability (e.g., ordinances on wages, developing a medium term expenditure framework to be eventually passed by parliament, etc.). On one hand, coordinating with IMF seems like a positive and desired feature of multilateral action. On the other, lack of significant reforms such as civil service reform or strengthening the institutions related to enforcing the fiscal rule could be seen as a missed opportunity for the Bank. Nevertheless the ICR and ICRR rated the operation as satisfactory.

5.13 In Tunisia and Cote d'Ivoire, the operations (both standalone DPOs) took place as emergency operations aimed to provide budget support to the interim government and in a post conflict situation respectively. Both countries lacked strong institutions and track records, and needed support to carry out social sector programs (in Tunisia) and reestablish economic functions (Cote d'Ivoire). In the case of Tunisia, the budget support was not accompanied with a proper sustainability analysis. Additionally, no prior actions on the fiscal side were selected, and the fact that there was no IMF program in place is worrisome in terms of fiscal sustainability after the loans have been disbursed and spent. A programmatic series in Tunisia seems better designed, but the implementation might not have been credible because the interim government lacked capacity to commit to medium-term reforms. Similarly, in Cote d'Ivoire, a standalone DPO without transformative prior actions could not be expected to have deep institutional impact on the macro-fiscal side. Even though in this case (similar to Romania) many structural reforms were left to the IMF program to deal with, the Bank's PAs selected did have not sufficient strength to advance significant reforms. The ICR and ICRR rated the operation in Cote d'Ivoire as moderately satisfactory.

5.14 Next is a group of DPOs with mixed fiscal and structural objectives. This may make it difficult to formulate consistent macro-fiscal programs as was the case with Benin, for example. On the other hand, when there are strong institutions and track records (e.g., Peru), such a DPO can be well designed and implemented.

Table 5.3.List and characteristics of DPOs in group 2 (mixed objectives—macro-fiscal and structural--DPOs)

II.	DPO's with some fiscal objectives to improve public finances as well as other structural reforms to accelerate growth			IMF Program Present
DPO		Macro/Fiscal Measures supported by Prior Actions		
	DPL rationale and clarity of macro diagnostics	Consistency	Institutional depth	
Indonesia	Poor infrastructure has become a key growth constraint. As a result, DPL series targeted infrastructure reforms as well as reforms to promote private sector development. On the macro side the operation supported preventive steps to confront the 2008 global crisis in terms of tightening regulations in the financial and insurance markets.	During the period of the series the macro-economic policy stance was strong and credible. The Bank and the IMF (Article IV) presenting a coherent and credible framework in support of the DPL.	No PA on the Macro/fiscal side were included in the series.	Yes
Benin	The Benin operation focus on private sector development, infrastructure, and progress towards MDGs. Macro and fiscal issues were not clearly identified, while the IMF had active 3 year PRGF in place during the operation. The operation did not have sufficient diagnostic work, preventing the Bank from identifying and addressing key development constraints.	The discussion ofmacro issues is limited and focused on the next two years' fiscal and current account deficit imbalances. There is no discussion of structural reform on the revenue and expenditure side that may contribute to fiscal sustainability.	The IMF has raised the issues of the increasing wage bill and the need for civil service reforms in order to contain it. The Fund expected the Bank to take the lead in this area. However, the series does not include PAs addressing this area. Most of the PAs in the fiscal areas are related to governance such as audits and procurement. There is a lack of structural depth in PAs selected.	Yes
St. Lucia	The PAD identified well the main issues and sources of vulnerability (e.g., external shocks, natural disasters, as well as high dependence on FDI and tourism. Despite relatively benign macro situation, measures were needed strengthen public finances. The operation featured many objectives and policy areas including governance, tax issues, PSD, and safety nets.	The operation supported specific improvements in public finances (e.g., merging of certain ministries, tax policy (VAT) and customs reforms. The reforms supported are in line with the macro diagnostics. However, there is a disconnect between the short-term time horizon of the operation and the long-time time it takes to implement these reforms.	Half of the PAs are focused on strengthening public finances. Capping public employment could have relied on a more long term civil service reform. As presented these measures could be easily be reversed. The VAT and customs reforms seem strong and institutionally relevant.	Yes
Peru	The series has a comprehensive discussion of the macro and debt situation as it evolves over time and encounters the 2008-09 global financial crisis. PD shows the government's good track record with macroeconomic and fiscal management with flexibility to deal with external shocks. The series focuses on sustainability and transparency of public finances and improving efficiency. Some vulnerabilities such as the currency composition of debt and the dependency on natural resources (and, therefore, vulnerabilities to international prices and exchange rates) are raised in the macro fiscal framework.	Large part of the series corresponds to a contingent DDO (70% of the series) in case of further deterioration of the external situation. This operation shows a sound use of this type of instruments, assisting the country by signaling to international markets the country's solid macro fiscal management and a good track record of fiscal discipline. The operation supported important reforms at the same time as shielding it from external shocks.	The PAs on the fiscal side seem strong and institutionally transformative. They follow a logical sequence over the series and are well monitored. Strengthening the fiscal rule, improveming the tax regime and regional resource allocations to bridge spacial inequities. As the series evolves, PAs become more specific reaching local governments.	Yes

Source: IEG team review.

5.15 In terms of the adequacy of the macro-fiscal framework in DPOs that had both fiscal and sectoral reforms supported, three out of four operations (Indonesia, St. Lucia, and Peru) presented a consistent and strong discussion of the macro situation. They tackle how the global crisis generated new challenges in terms of fiscal sustainability and presented consistent and credible analysis. Benin, on the other hand, did not present a strong and credible macro framework, in particular due to the short-term nature of the analysis focusing only on two years. Additionally, the program document of the Benin operation did not discuss some of the most important issues affecting fiscal sustainability, such as civil service reform, which was identified by IMF. Also, three operations did include a number of fiscal prior actions and triggers in support of the

DPL (Peru, St. Lucia, and Benin), while Indonesia did not include a particular PA on the fiscal side.

5.16 While these three countries included fiscal-related PA, institutional depth of the supporting prior action varies significantly between these operations.

5.17 Peru presents a good practice. A strong program, with a well-articulated macro framework, listing in a logical manner all the reform steps needed to strengthen the fiscal position. It must be acknowledged that Peru started from a very robust fiscal position, good track record, and that the country had healthy institutions in place. The program in Peru helped further develop existing institutions and most of the loan was of precautionary nature, helping Peru signal to the international financial markets its policy resolve to help it weather the effects of the crisis. The fact that the Peru operation was a series helped lengthen the dialogue with the country, permitting flexibility in the face of shocks. On the other hand, in the St. Lucia case, a standalone operation, there was a disconnect between the long-term nature of the reforms and the short term horizon of the loan. Even though St. Lucia's program document presented a thorough macro framework and was candid about the main vulnerabilities the country is subject to (e.g., natural disasters, high dependency on tourism, and foreign direct investment), the prior actions selected did not address these vulnerabilities and the reform program supported was unable to succeed in the time frame as some of the risks materialized (i.e., Hurricane Tomas). Additionally, some of the measures supported by the operation did not have enough institutional depth and could be easily reversed (e.g., civil service measures). Similarly in Benin, the measures supported were easily reversible (e.g., audits and procurement measures).

Table 5.4.List and characteristics of DPOs in group 3 (mainly sectoral DPOs)

III.	**Sectoral focused DPO's**			**IMF Program Present**
DPO		Macro/Fiscal Measures supported by Prior Actions		
	DPL rationale and clarity of macro diagnostics	**Consistency**	**Institutional depth**	
Albania	The operation was designed in an environment of major macroeconomic risks and no IMF program. By not including explicit actions to contain the fiscal deficit and the overall debt level, risks mitigation measures were insufficient, and the risks materialized. The loan had two objectives: provide some countercyclical resources for the government to confront the global and European crises and improve the social safety nets and health sector spending.	The macro discussion acknowledges that the government identified in the letter of development policy many uncertainties for the future that may emerge from the global and European crises. However, the operation does not identify any fiscal action that could strengthen the fiscal stance to shield better the country against future shocks. There is a question why the IMF was not present with a program.	No PA addresses strengthening public finances given the uncertainties about the future. It is unclear why the Bank proceeded with the DPL without the Fund program because the fund could have, in principle, provided contingent resources to face the future shocks.	No
Poland	The operation is narrowly focused on reforms to achieve energy efficiency. The macro discussion in the PAD, however, sound. It acknowledges deep institutional reforms, including the introduction of fiscal rules. The debt sustainability analysis is thorough.	The PAD acknowledges that the institutional framework in place assures sustainability.	All PAs are focused on reforms aimed at enhancing nergy efficiency.	Yes
Latvia	The objective of the series is to contribute financing to the large financial package from the IMF and the EC to help Latvia face the global crisis. (IMF and the EC were the key financiers of the macro program)	The DPL focused on mitigating the cost of fiscal contraction on the safety nets and on maintaining access to health and education.	The PA actions did not address macro/fiscal issues. These issues are addressed by the IMF and EC. The PA focus, instead, on making sure that basic services are being funded.	Yes
Mexico	The program document shows the strong macroeconomic performance prior to the global financial crisis. Mexico shows solid growth with low inflation; a steady reduction of public debt, and strengthened corporate balance sheets; limited current account deficit; and a profitable and well capitalized banking sector. This had been underpinned by a highly credible and strong policy framework, including a successful inflation targeting regime that had supported the commitment to the flexible exchange rate; a rules-based fiscal framework; and strong and sophisticated financial sector supervision. The financial crisis posed challenges, however. The authorities took robust and timely measures to respond to the deteriorating global situation. This included securig contingent financing from the IMF (FCL).	The macro and fiscal stability were mostly addressed by tye IMF program and the IMF FCL. The public debt was expected to remain within manageable proportions despite a significant increase in the debt-to-GDP ratio until 2011. The IMF assessed the 2010 budget framework as appropriate. Overall, the macro framework is considered to be consistent and credible.	The DPL did not include specific macro conditions, prior actions or indicators, even though the first part of its Program Development Objective was to, "support stimulus of the economy." The IMF had a program and the FCL. Also, Economic Policies in response to the Global Crisis DPL, approved a month after this DPL, included macro-related conditions, prior actions and indicators.	Yes

Source: IEG team review.

5.18 The last group of DPOs corresponds solely to sectoral policy loans. In this group, two operations (Poland and Mexico) had a strong macro framework, and well organized discussion of the fiscal standing of the country. The other two operations (Albania and Latvia) however, presented a weaker fiscal analysis. In Albania, the operation took place in an environment of major macroeconomic risks and no IMF program. It is, therefore, surprising that the DPO did not include explicit actions to contain the fiscal deficit and the overall debt level, and risks mitigation measures were insufficient and the risks finally did materialized. The standalone nature of the operation might have contributed to the lack of further and in-depth analysis needed to account for the possible fiscal pitfalls encountered. The rating of ICRR for Albania was unsatisfactory (a significant downgrade from the satisfactory rating of the ICR). The IEG evaluation argued that:

"While the project was responsive to country conditions and Bank strategy, its design did not correspond well to the program development objectives (PDO). The choice of lending instrument and implementation time frame were inadequate for achieving the objectives. The results framework would offer little guidance on what improvements in outcomes could be expected, and in particular how the outcome indicators chosen would meet the PDO" (ICRR p. 7).

5.19 In Latvia, even though macro and fiscal consideration were highly relevant in this operation, the objective of the operation was contributing with resources mainly to the social safety net program. In terms of fiscal and macro stability measures, the main players were IMF and the European Commission. The operation did not incorporate prior action sin the macro-fiscal realm. Nevertheless the operation rating in both the ICR and ICRR was satisfactory.

5.20 Based on these case studies and the previous portfolio and consistency analyses, it is possible to draw some tentative examples of good practices and lessons (Box 5.1).

Macro frameworks in standalone vs. programmatic series DPOs

5.21 There are other meaningful groupings that can shed light on some commonalities between different types of budget support operations worth analyzing. Some important differences in the macro-fiscal framework and overall robustness of the DPO were found in standalone or single tranche versus programmatic operations.

5.22 There are two clear difference between these two types of budget support operations when it comes to the macro-fiscal framework. First, the length of the operation (standalone or programmatic) affects the capacity to link substantive prior actions to significant, longer-term fiscal and macroeconomic stability results. Second, the time spent in preparing the operation (e.g., emergency operations) could be affecting the dialogue with the borrower country as well as the capacity of the Bank staff to dig deeper into relevant issues regarding macro and fiscal sustainability. Third, standalone operations tend to show some specific weaknesses.

Box 5.1. Good practices in macro/fiscal framework from the case studies

- Macro-fiscal objectives should have clearly articulated measures that are realistic and tailored to the problem at hand. Whether or not these measures should be prior actions is a matter of judgment and in relation to the related IMF program.
- A strong track record is important. Good track record of prudent fiscal management and sustainable macro policies can be linked to a higher credibility of the macro framework. It also provides room for strengthening further the already in place institutions. Thus advancing further the fiscal sanding of the borrower country and providing the proper signaling to the international and internal financial markets.
- Collaboration with the IMF is important. When presenting a credible and consistent macro framework, most operations relied on the Fund's Article IV consultation and debt sustainability analysis. Choosing prior actions identified by the IMF could also be a good practice that could keep consistency between the two multilaterals.
- If track record is poor and there is no IMF program, this should be a "yellow flag" in reviewing the quality of the macro-fiscal program. In those cases, the burden of proof increases on the Bank to demonstrate that macro-fiscal program is sufficiently articulated, consistent, credible and sustainable, and operational reviews should scrutinize those operations with greater attention.
- Accompanying the DPO with sufficient diagnostic work, NLTA and AAA (Public Expenditure Reviews) seems to pay off. Stronger macro-fiscal frameworks were found in countries with more diagnostic work prior to the operation.
- Adopting focused and realistic fiscal objectives can help improve macro-fiscal frameworks, design of prior actions and results. When the objectives are more precise and less ambitious, the likelihood of success improves.
- High level of borrower and Bank ownership and significant attention to the quality of policy dialogue is needed to increase the probability of success.

Source: IEG.

5.23 Some of the challenges of the standalone or single tranche operations reviewed are listed below.

- In standalone operations, there is sometimes a disconnect between objectives and prior actions. This may be a matter of substance or timing. Significant reforms with institutional transformative potential take time. As a result, standalone operation might not be well suited to follow through on these kinds of reforms. (Yet, they may open the door and dialogue leading to the deeper, multi-year engagement). For deeper policy reforms, a programmatic series could help the Bank and the borrower country engage in a broader dialogue, longer-term reforms, and make more sustained adjustments to achieve the reforms objectives. This may also result in longer-term building of institutions and development of trust among the Bank and the country authorities.

- The standalone operations may be particularly useful in addressing narrower short-to-medium term or sectoral objectives and in reengagement, post-conflict, uncertain, and crisis situations.

In many cases, programmatic series are also accompanied by NLTA and AAA. This might help both the Bank and the Government further understand the development challenges and adjust the operation according to the results of the analytical work.

Box 5.2. Weaknesses in macro/fiscal framework from case studies

- Conflicting macro-objectives. In some cases fiscal expenditure needs to be contracted to reach macro balances, but at the same time the operation is supporting expansive social policies without paying enough attention on the sustainability of the programs.
- Inadequate accounting of external vulnerabilities and political economy risks specifically tailored to policy situations at hand.
- Insufficiently articulated sustainability analysis.
- Weaker macro frameworks in sectoral operations and those without IMF programs.

Source: IEG team review in this paper.

5.24 The following seven lessons can be drawn from the reviewed case studies:

- Good-practice macro-fiscal frameworks feature close correspondence between macro-fiscal objectives and the articulation of fiscal measures.
- The government's track record can significantly influence the realism of the design and success of the DPOs.
- When the track record and IMF program are absent, extra time, effort and scrutiny is needed in preparation and review of DPOs.
- Well defined and attainable objectives within the timeframe of the operation can significantly influence the quality and implementation of macro-fiscal frameworks.
- Standalone DPOs may be usefully combined with technical assistance to assure sustainability of reforms and provide the borrower with further analysis on the reforms and their possible fiscal and macro effect in the medium and long run.
- Political economy risks could usefully be explained in the particular country and policy (not generic) context and linked to specific macro-fiscal reform program.

6. Correlates of DPO outcomes and quality and implementation of macro-fiscal frameworks: A preliminary statistical analysis

6.1 This section provides complementary, preliminary statistical evidence on the role of macro-fiscal frameworks. It seeks to empirically explore two questions.

6.2 First, it aims to identify correlates of Independent Evaluation Group (IEG) ratings as measures of the performance of Development Policy Operations (DPOs) that may be related to the quality of the macro-fiscal frameworks as defined in this paper, including elements of consistency, track record, credibility, and sustainability. To this end, an index of consistency described above is combined with other indices into a composite index of the quality of the design of macro-fiscal framework. IEG DPO ratings are regressed on this index (and its components) and other project, country, and external independent variables. Also developed was an index of the implementation of the fiscal framework and used in specific regressions.

6.3 Second, the analysis aims to identify the correlates of the quality of the macro-framework itself using a set of regressors, and to explore the links between the quality of design and the implementation of the fiscal frameworks.

6.4 The analysis uses the data on DPOs that are closed and with IEG Implementation Completion Report (ICR) reviews completed in the period 2004-13. All the data and variable definitions, as well detailed statistical tables are provided in the annex. In this summary, we highlight four main, preliminary results. They appear broadly consistent with findings from the findings using other methods discussed above, including the elements of the portfolio analysis, descriptive statistics of the consistency index, and insights and lessons from the case studies.

6.5 First, summary statistics for IEG ratings of DPOs, and indices of quality of macro-fiscal frameworks and the implementation of macro-fiscal frameworks suggest that (i) IEG ratings of DPOs slightly worsened in the past three years compared to the whole period of analysis, but (ii) the quality of design and implementation of macro-fiscal frameworks improved in recent years. It is noted that the FY11-13 only include a relatively small number of DPOs (closed and with IEG ICR ratings) compared with the rest of the sample, which should caveat the first statistic (Table 6.1).

Table 6.1 Comparison of IEG ratings, Macro design quality, and macro implementation quality before and after 2011

	IEG rating		Design quality		Implementation quality	
	MEAN	ST. DEV.	MEAN	ST. DEV.	MEAN	ST. DEV.
2004-2011	3.23	0.93	-0.04	0.60	-0.03	0.73
2011-2013	3	0.94	0.24	0.73	0.15	0.76
Test of difference (*p*-value)	0.07		0.04		0.01	

Note: Mann-Whitney-Wilcoxon test for IEG rating. Kolmogorov-Smirnov test for design and implementation quality. The last row shows the p-value from the test that compares the distribution of each variable before and after 2011. A common test of distributions is the Kolmogorov-Smirnov test, which compares the location (mean) and the shape (standard deviation) of two distributions. The null hypothesis is no difference. This test is appropriate when the distributions are approximately normal. Since the IEG ratings are not normally distributed, a Mann-Whitney-Wilcoxon test was used, which is more appropriate in this situation. The null hypothesis is again no difference in the distributions. Whatever the test, the last row shows that the difference in the distribution of the variable before and after 2011 is statistically significant (at $p<.1$ for IEG ratings, and at $p<0.04$ and $p<0.01$ for the macro design and implementation quality, respectively).

6.6 Second, the elements of quality of the macro frameworks design are positively correlated with the IEG ratings. In particular, ceteris-paribus, measure of track record ("backward-looking credibility") and the coverage of quasi-fiscal risks are statistically significantly correlated with the IEG ratings at less than $p<0.1$ (Table 6.2). The first five columns show the correlation between the five different aspects of the macro design quality and IEG ratings, while controlling for other important factors. The last column shows the correlation between the combined index of these five measures and the IEG ratings. Each of the five variables is standardized to mean zero and standard deviation one. The combined index is a simple average of the five measures (therefore also having a mean of zero). The coefficients for each of the variables shows the effect of a one standard deviation increase. For example, a one standard deviation increase in the backward-looking credibility is associated with a .1 point increase in the IEG rating, on a scale from 1 (highly unsatisfactory) to 6 (highly satisfactory). Of interest is also that programmatic DPOs and DPOs with economic policy thematic focus are positively correlated with IEG ratings, which in some regressions also show statistical significance.

Table 6.2 Correlation between IEG ratings and macro-design quality

Dependent variable is IEG rating	(1) BW cred.	(2) FW cred.	(3) QF risks	(4) DSA	(5) Cons.	(6) Combined
KEY VARIABLES						
Backward-looking credibility	0.102*					
	(0.055)					
Forward-looking credibility		0.107				
		(0.098)				
Quasi-fiscal risks coverage			0.142**			
			(0.063)			
Debt sustainability coverage				-0.032		
				(0.067)		
Macro design consistency					0.051	
					(0.071)	
Combined macro design quality index						0.157
						(0.142)
PROJECT-LEVEL VARIABLES						
TTL previous projects	0.006	0.028	-0.017	-0.018	-0.030	-0.032
	(0.081)	(0.066)	(0.057)	(0.057)	(0.070)	(0.057)
DPO size	0.111	0.095	0.134	0.129	0.050	0.112
	(0.126)	(0.114)	(0.091)	(0.094)	(0.115)	(0.098)
Programmatic DPO	0.166	0.228	0.282**	0.286**	0.175	0.274*
	(0.205)	(0.175)	(0.140)	(0.139)	(0.175)	(0.142)
Econ. sector	0.201	0.280	0.244*	0.293**	0.300	0.220
	(0.224)	(0.170)	(0.141)	(0.144)	(0.185)	(0.146)
COUNTRY-LEVEL VARIABLES						
# previous DPOs	-0.005	-0.005	0.003	0.000	-0.002	-0.000
	(0.016)	(0.014)	(0.011)	(0.011)	(0.014)	(0.011)
IMF program overlap	0.016	0.012	0.251*	0.221	0.130	0.227
	(0.206)	(0.180)	(0.150)	(0.147)	(0.183)	(0.148)
GDP per capita (PPP)	-0.129	0.065	0.013	0.062	0.082	0.048
	(0.267)	(0.176)	(0.126)	(0.127)	(0.166)	(0.126)
GDP (PPP)	-0.077	-0.072	-0.122	-0.109	-0.033	-0.082
	(0.134)	(0.117)	(0.093)	(0.103)	(0.121)	(0.104)
Africa region	-0.420	-0.112	-0.198	-0.206	-0.095	-0.106
	(0.362)	(0.329)	(0.229)	(0.234)	(0.290)	(0.244)
CPIA score	0.105	0.056	0.222	0.122	0.134	0.179
	(0.305)	(0.235)	(0.207)	(0.202)	(0.226)	(0.202)
EXTERNAL SHOCKS VARIABLES						
TOT change (avg. prev. 3 years)	-0.622	-1.681	-1.130	-0.904	-1.358	-1.170
	(1.410)	(1.288)	(1.140)	(1.146)	(1.302)	(1.139)
GDP growth (avg. prev. 3 years)	-0.003	0.012	0.009	0.012	-0.004	0.013
	(0.038)	(0.033)	(0.024)	(0.024)	(0.026)	(0.025)
Constant	3.723*	2.191	2.453*	2.192	1.718	1.749
	(2.178)	(1.776)	(1.402)	(1.434)	(1.742)	(1.458)
Observations	132	192	256	260	174	255
Adjusted R-squared	-0.028	0.017	0.066	0.044	0.007	0.041

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

6.7 Third, the quality of the macro frameworks implementation is not statistically strongly correlated with the IEG ratings. However, after controlling for macro implementation quality (2nd column), the macro design quality is statistically significantly correlated with the IEG ratings (table 6.3 below), unlike in the previous table above (table 6.2). The macro implementation quality is standardized to mean zero and standard deviation one. The first column examines the correlation without taking into account the quality of the macro design. The second column takes the design into account. The coefficients for the macro implementation variable show the effect of a one standard deviation increase.

Table 6.3 Correlation between IEG rating and macro implementation quality

Dependent variable is IEG rating	(1) Implementation	(2) Controlling for design quality
KEY VARIABLES		
Macro implementation quality	-0.074	-0.244
	(0.154)	(0.196)
Combined macro design quality index		0.397**
		(0.193)
PROJECT-LEVEL VARIABLES		
TTL previous projects	-0.027	-0.022
	(0.073)	(0.073)
DPO size	0.052	0.059
	(0.116)	(0.113)
Programmatic DPO	0.185	0.195
	(0.177)	(0.176)
Econ. sector	0.351*	0.244
	(0.187)	(0.184)
# previous DPOs	-0.001	0.001
	(0.014)	(0.014)
COUNTRY-LEVEL VARIABLES		
IMF program overlap	0.110	0.114
	(0.190)	(0.191)
GDP per capita (PPP)	0.089	0.062
	(0.173)	(0.168)
GDP (PPP)	-0.035	-0.023
	(0.121)	(0.121)
Africa region	-0.116	-0.011
	(0.282)	(0.296)
CPIA score	0.107	0.159
	(0.230)	(0.216)
EXTERNAL SHOCKS VARIABLES		
TOT change (avg. prev. 3 years)	-1.432	-1.670
	(1.316)	(1.338)
GDP growth (avg. prev. 3 years)	-0.006	0.001
	(0.027)	(0.025)
Constant	1.759	1.297
	(1.714)	(1.840)
Observations	172	172
Adjusted R-squared	0.005	0.030

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
Outcome is IEG rating. Standard errors clustered by country-programmatic series.

6.8 Fourth, the table shows that the quality matters for implementation: macro frameworks design is generally positively correlated with the macro implementation quality (all but the macro design consistency measure) (table 6.4). The macro implementation quality is standardized to mean zero and standard deviation one. Each of the five variables is standardized to mean zero and standard deviation one. The combined index is a simple average of the five measures (therefore also having a mean of zero). First five columns show the correlation between five difference aspects of the macro design quality and the macro implementation quality, while controlling for other important factors. The last column shows the correlation between the combined index of these four measures and the macro implementation quality. The coefficients for each of the variables shows the effect of a one standard deviation increase. For example, a one standard deviation increase in the backward-looking credibility is associated with a .26 standard deviation increase in macro implementation quality.

Table 6.4 Correlation between macro implementation quality and macro design quality

Outcome is Macro implementation quality	(1) BW cred.	(2) FW cred.	(3) QF risks	(4) DSA	(5) Cons.	(6) Combined
Key variables						
Backward-looking credibility	0.263**					
	(0.103)					
Forward-looking credibility		0.526*				
		(0.274)				
Quasi-fiscal risks coverage			0.175**			
			(0.067)			
Debt sustainability coverage				0.180***		
				(0.055)		
Macro design consistency					0.032	
					(0.051)	
Combined macro design quality index						0.560***
						(0.093)
Project-level variables						
TTL previous projects	0.021	0.045	0.035	0.051	0.030	0.039
	(0.042)	(0.047)	(0.045)	(0.044)	(0.046)	(0.039)
DPO size	-0.059	-0.024	-0.025	-0.047	-0.042	-0.029
	(0.098)	(0.069)	(0.066)	(0.064)	(0.068)	(0.057)
Programmatic DPO	-0.107	0.026	0.086	0.132	0.121	0.095
	(0.132)	(0.125)	(0.111)	(0.108)	(0.112)	(0.097)
Econ. Sector Refer	0.183	0.231**	0.307***	0.216**	0.301***	0.110
	(0.123)	(0.114)	(0.103)	(0.103)	(0.102)	(0.105)
Country-level variables						
# previous DPOs	0.008	0.014	0.023**	0.014*	0.016*	0.017**
	(0.010)	(0.010)	(0.009)	(0.008)	(0.008)	(0.008)
IMF program overlap	-0.085	-0.232*	-0.258**	-0.314***	-0.288**	-0.233**
	(0.115)	(0.119)	(0.113)	(0.114)	(0.115)	(0.101)

GDP per capita (PPP)	0.202	0.062	0.050	0.147	0.129	0.090
	(0.222)	(0.165)	(0.125)	(0.122)	(0.128)	(0.116)
GDP (PPP)	0.075	0.065	0.007	0.083*	0.043	0.052
	(0.063)	(0.048)	(0.053)	(0.048)	(0.054)	(0.041)
Africa region	0.113	0.054	-0.113	0.038	-0.060	0.110
	(0.216)	(0.182)	(0.159)	(0.148)	(0.170)	(0.132)
CPIA score	-0.075	-0.102	-0.072	-0.284**	-0.230*	-0.112
	(0.130)	(0.145)	(0.131)	(0.122)	(0.124)	(0.109)
External shocks variables						
TOT change (avg. prev. 3 years)	-0.916	-1.275*	-1.142	-0.799	-0.641	-0.988
	(0.897)	(0.738)	(0.794)	(0.787)	(0.781)	(0.642)
GDP growth (avg. prev. 3 years)	0.007	0.011	0.007	0.019	0.009	0.018
	(0.021)	(0.022)	(0.019)	(0.019)	(0.019)	(0.016)
Constant	-2.399	-1.664	-0.220	-1.741	-0.818	-1.459
	(1.714)	(1.286)	(1.148)	(1.185)	(1.202)	(0.963)
Observations	123	153	173	174	174	174
Adjusted R-squared	0.075	0.085	0.159	0.139	0.090	0.235

Robust standard errors in parentheses.

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

Outcome is implementation quality. Standard errors clustered by country-programmatic series.

References

Alesina, Alberto, and Roberto Perotti. 1995. Fiscal expansions and fiscal adjustments in OECD countries. No. W5214. National Bureau of Economic Research.

Alesina, Alberto, and Guido Tabellini. 1989. "External debt, capital flight and political risk." Journal of International Economics 27, no. 3: 199-220.

Bogetic Zeljko and Arye Hillman eds. 1995. Political Economy of Tax Bases, Tax Policies, and Tax Evasion. World Bank Regional and Sectoral Studies, World Bank.

Bogetic, Zeljko, Karlis Smits, Nina Budina, and Sweder van Wijnbergen 2010. Long-Term Fiscal Sustainability and Risks in an Oil-Rich Country: The Case of Russia, World Bank Policy Research Paper Series No. 5240 (March).

Denizer, Cevdet, Daniel Kaufmann, and Aart Kraay. 2011. Good Countries or Good Projects? Macro and Micro Correlates of World Bank Project Performance. World Bank Policy Research Working Paper 5646.

Dobronogov, Anton, Alan Gelb and Fernando Brant Saldanha. 2014. How Should Donors Respond to Resource Windfalls in Poor Countries: From Aid to Insurance. World Bank Policy Research Paper Series No. 6952.

Fischer, Stanley, Ratna Sahay, and Carlos A. Végh. 2002. Modern hyper-and high inflations. No. W8930. National Bureau of Economic Research.

Geli, Patricia, Aart Kraay, and Hoveida Nobakht. 2014. Predicting World Bank Project Outcome Ratings. World Bank Policy Research Working Paper 7001.

Geli, Patricia, Peter Moll, and Pablo Saavedra, 2015. Correlates of Success of Development Policy operations, World Bank Policy Research paper.

Gupta, Sanjeev, Benedict Clements, Emanuele Baldacci, and Carlos Mulas-Granados. 2005. "Fiscal policy, expenditure composition, and growth in low-income countries." Journal of International Money and Finance 24, no. 3: 441-463.

Gupta, Sanjeev, Marijn Verhoeven, and Erwin Tiongson. 1999. Does higher government spending buy better results in education and health care? International Monetary Fund.

Hillman, Arye L. 2003. Public Finance And Public Policy: Responsibilities And Limitations Of Government. Cambridge University Press.

IEG. 2010. Poverty Reduction Support Credits: An Evaluation of World Bank Support. Independent Evaluation Group, The World Bank.

IEG. 2012. Designing a Results Framework for Achieving Results: A How to Guide.

IEG. 2010. Poverty Reduction Support Credits: An Evaluation of World Bank Support.

IEG. 2010. PRSC Results Framework Evaluation.

IEG. 2013. Guidelines for Reviewing World Bank Implementation Completion and Results Reports.

IEO 2005. Evaluation of Structural Conditionality in IMF-Supported Programs, Issues paper for an Evaluation by the Independent Evaluation Office of the International Monetary Fund.

IEO 2013. Fiscal Adjustment in IMF-Supported Programs: Revisting the 2003 IEO Evaluation. Evaluation Update, Independent Evaluation Office of the International Monetary Fund.

IEG. 2015a. The Role of Knowledge on Public Expenditures and Resource Allocation in Development Policy Operations (DPOs), IEG learning product (May).

IEG. 2015b. Quality of Results Frameworks in Development Policy Operations (DPOs), IEG learning product (May).

IMF. 2003. Signalling Assessments of Member Policies. Prepared by the Polidy Development and Review Department, January 8.

IMF. 2010. Implementation of the Joint Management Action Plan on Bank-Fund Collaboration, prepared by Bank's Poverty Reduction and Economic Management Network and the Fund's Strategy, Policy, and Review Department, March 3.

IMF and World Bank. 2011. Guidance Note on Good Practice in Annual Bank-Fund Team Collaboration Reviews.

Moll, Peter, Patricia Geli and Pablo Saavedra (2015). "Correlates of Success of Development Policy Operations," World Bank Policy Research Paper Series No. 2015/01/28.

OP/BP 8.60. Development Policy Financing (updated July 2014).

Ossowsky, Rolando 2014. Extractive Industries Sector: Essentials for Economists and Public Finance Professionals; Macro-Fiscal Frameworks in Resource-Rich Countries, (November 10), draft.

OPCS 2009. DPL Training for IEG Staff.

OPCS 2012. DPO 2012 Development Policy Lending Retrospective: Results, Risks, And Reforms. Operations Policy and Country Services, the World Bank, October 2012.

OPCS 2013. Guidance Note for Development Policy Operations: Macroeconomic Policy Frameworks and Collaboration with the IMF in Development Policy Operations. September 2013.

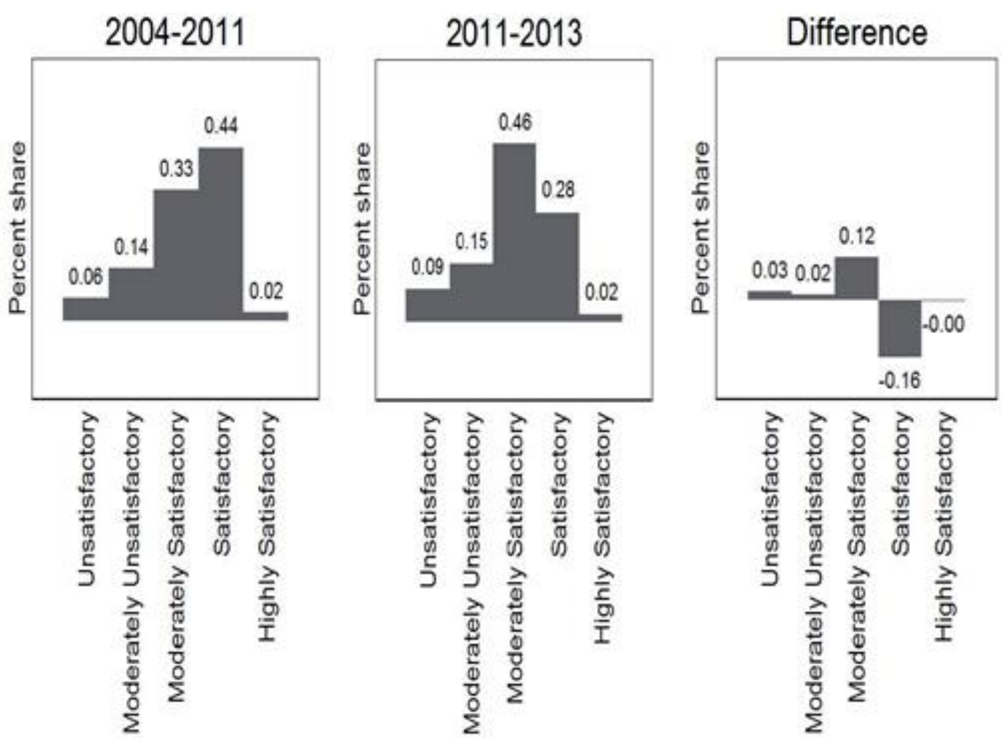
OP8.60. 2014. Operational Policies. February 2012, revised July 2014, World Bank, Washington D.C.

Smets, Lodewijk, and Stephen Knack. 2014. World Bank Lending and the Quality of Economic Policy. World Bank Policy Research Working Paper 6924.

Vegh, Carlos 1990. Credibility and the Dynamics of Stabilization Policy, IMF Working Paper No. 90/110.

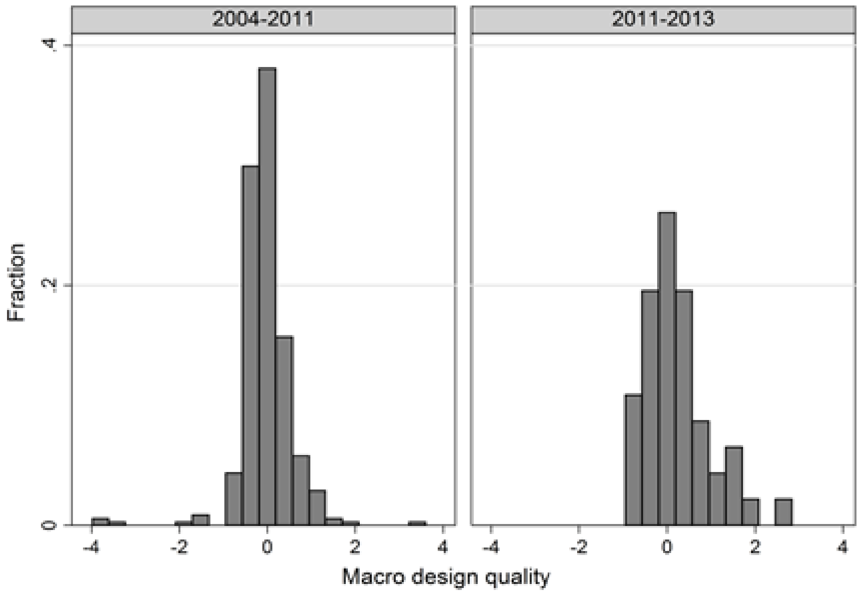
Appendix A. Statistical Annex

Figure A-1. IEG Ratings before and after 2011



IEG ratings on average lower in the period after 2011 (middle panel) than before 2011 (left panel). The right-most panel shows the difference for each rating category. Note that there were no "highly unsatisfactory" DPOs during the entire period 2004-2013.

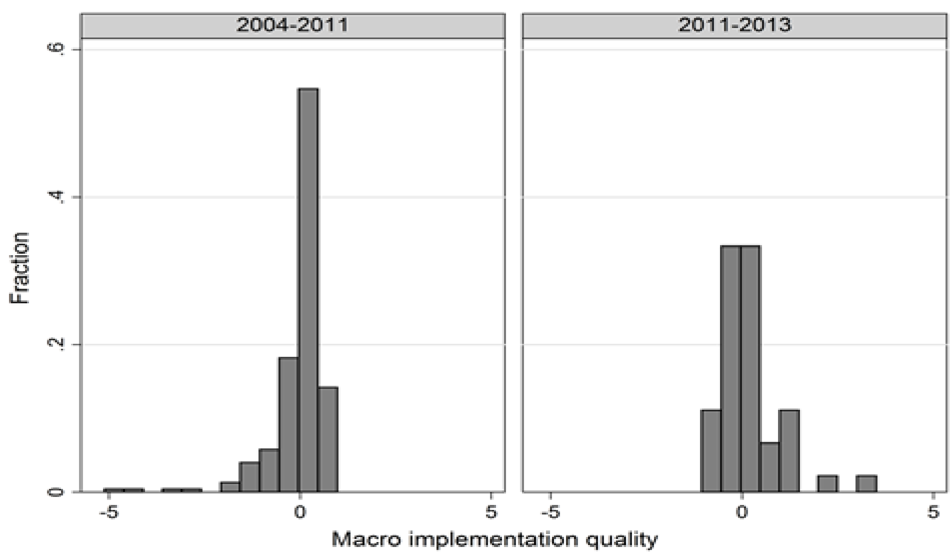
Figure A-2a Macro design quality before and after 2011



Quality of the design of macro frameworks improved after 2011 (right panel) relative to before 2011 (left panel). The macro design quality is a standardized measure with mean zero and standard deviation

The y-axis represents the percent share of data in each bar.

Figure A-2b Macro implementation quality before and after 2011



Implementation quality also improved after 2011 (right panel) relative to before 2011.(left panel).

Table A-1. Comparison of IEG ratings, Macro design quality, and macro implementation quality before and after 2011

	IEG rating		Design quality		Implementation quality	
	MEAN	ST. DEV.	MEAN	ST. DEV.	MEAN	ST. DEV.
2004-2011	3.23	0.93	-0.04	0.60	-0.03	0.73
2011-2013	3	0.94	0.24	0.73	0.15	0.76
Test of difference (*p*-value)	0.07		0.04		0.01	

Note: Mann-Whitney-Wilcoxon test for IEG rating. Kolmogorov-Smirnov test for design and implementation quality.

The table shows more formally what was presented graphically. The last row shows the p-value from the test that compares the distribution of each variable before and after 2011. A common test of distributions is the Kolmogorov-Smirnov test, which compares the location (mean) and the shape (standard deviation) of two distributions. The null hypothesis is no difference. This test is appropriate when the distributions are approximately normal. Since the IEG ratings are not normally distributed, a Mann-Whitney-Wilcoxon test was used, which is more appropriate in this situation. The null hypothesis is again no difference in the distributions. Whatever the test, the last row shows that the difference in the distribution of the variable before and after 2011 is statistically significant (at p<.1 for IEG ratings, and at p<0.04 and p<0.01 for the macro design and implementation quality, respectively).

Table A-2. Correlation between IEG ratings and macro-design quality

Dependent variable is IEG rating	(1) BW cred.	(2) FW cred.	(3) QF risks	(4) DSA	(5) Cons.	(6) Combined
Key variables						
Backward-looking credibility	0.102*					
	(0.055)					
Forward-looking credibility		0.107				
		(0.098)				
Quasi-fiscal risks coverage			0.142**			
			(0.063)			
Debt sustainability coverage				-0.032		
				(0.067)		
Macro design consistency					0.051	
					(0.071)	
Combined macro design quality index						0.157
						(0.142)
Project-level variables						
TTL previous projects	0.006	0.028	-0.017	-0.018	-0.030	-0.032
	(0.081)	(0.066)	(0.057)	(0.057)	(0.070)	(0.057)
DPO size	0.111	0.095	0.134	0.129	0.050	0.112
	(0.126)	(0.114)	(0.091)	(0.094)	(0.115)	(0.098)
Programmatic DPO	0.166	0.228	0.282**	0.286**	0.175	0.274*
	(0.205)	(0.175)	(0.140)	(0.139)	(0.175)	(0.142)
Econ. sector	0.201	0.280	0.244*	0.293**	0.300	0.220

	(0.224)	(0.170)	(0.141)	(0.144)	(0.185)	(0.146)
Country-level variables						
# previous DPOs	-0.005	-0.005	0.003	0.000	-0.002	-0.000
	(0.016)	(0.014)	(0.011)	(0.011)	(0.014)	(0.011)
IMF program overlap	0.016	0.012	0.251*	0.221	0.130	0.227
	(0.206)	(0.180)	(0.150)	(0.147)	(0.183)	(0.148)
GDP per capita (PPP)	-0.129	0.065	0.013	0.062	0.082	0.048
	(0.267)	(0.176)	(0.126)	(0.127)	(0.166)	(0.126)
GDP (PPP)	-0.077	-0.072	-0.122	-0.109	-0.033	-0.082
	(0.134)	(0.117)	(0.093)	(0.103)	(0.121)	(0.104)
Africa region	-0.420	-0.112	-0.198	-0.206	-0.095	-0.106
	(0.362)	(0.329)	(0.229)	(0.234)	(0.290)	(0.244)
CPIA score	0.105	0.056	0.222	0.122	0.134	0.179
	(0.305)	(0.235)	(0.207)	(0.202)	(0.226)	(0.202)
External shocks variables						
TOT change (avg. prev. 3 years)	-0.622	-1.681	-1.130	-0.904	-1.358	-1.170
	(1.410)	(1.288)	(1.140)	(1.146)	(1.302)	(1.139)
GDP growth (avg. prev. 3 years)	-0.003	0.012	0.009	0.012	-0.004	0.013
	(0.038)	(0.033)	(0.024)	(0.024)	(0.026)	(0.025)
Constant	3.723*	2.191	2.453*	2.192	1.718	1.749
	(2.178)	(1.776)	(1.402)	(1.434)	(1.742)	(1.458)
Observations	132	192	256	260	174	255
Adjusted R-squared	-0.028	0.017	0.066	0.044	0.007	0.041

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Outcome is IEG rating. Standard errors clustered by country-programmatic series.

- for description of variables, see at the bottom
- First five columns show the correlation between the five different aspects of the macro design quality and IEG ratings, while controlling for other important factors. The last column shows the correlation between the combined index of these five measures and the IEG ratings.
- Each of the five variables is standardized to mean zero and standard deviation one. The combined index is a simple average of the five measures (therefore also having a mean of zero).
- The table shows that the quality of the macroframeworks design is generally positively correlated with the IEG ratings. In particular, ceteris-paribus, backward-looking credibility and the coverage of quasi-fiscal risks are statistically significantly correlated with the IEG ratings at less than p<0.1
- The coefficients for each of the variables shows the effect of a one standard deviation increase. For example, a one standard deviation increase in the backward-looking credibility is associated with a .1 increase in the IEG rating, on a scale from 1 (unsatisfactory) to 5 (highly satisfactory).

Table A-3. Correlation between IEG rating and macro implementation quality

Dependent variable is IEG rating	(1) Implementation	(2) Controlling for design quality
Key variables		
Macro implementation quality	-0.074	-0.244
	(0.154)	(0.196)
Combined macro design quality index		0.397**
		(0.193)
Project-level variables		
TTL previous projects	-0.027	-0.022
	(0.073)	(0.073)
DPO size	0.052	0.059
	(0.116)	(0.113)
Programmatic DPO	0.185	0.195
	(0.177)	(0.176)
Econ. sector	0.351*	0.244
	(0.187)	(0.184)
# previous DPOs	-0.001	0.001
	(0.014)	(0.014)
Country-level variables		
IMF program overlap	0.110	0.114
	(0.190)	(0.191)
GDP per capita (PPP)	0.089	0.062
	(0.173)	(0.168)
GDP (PPP)	-0.035	-0.023
	(0.121)	(0.121)
Africa region	-0.116	-0.011
	(0.282)	(0.296)
CPIA score	0.107	0.159
	(0.230)	(0.216)
External shocks variables		
TOT change (avg. prev. 3 years)	-1.432	-1.670
	(1.316)	(1.338)
GDP growth (avg. prev. 3 years)	-0.006	0.001
	(0.027)	(0.025)
Constant	1.759	1.297
	(1.714)	(1.840)
Observations	172	172
Adjusted R-squared	0.005	0.030

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Outcome is IEG rating. Standard errors clustered by country-programmatic series.

- This table shows the correlation between the macro implementation quality and the IEG ratings, controlling for other important factors.

- The macro implementation quality is standardized to mean zero and standard deviation one.
- The first column examines the correlation without taking into account the quality of the macro design. The second column takes the design into account.
- Both columns suggest that the quality of the macroframeworks implementation is not statistically strongly correlated with the IEG ratings. Note that in the second column, after controlling for macro implementation quality, the macro design quality is statistically significantly correlated with the IEG ratings, unlike in the previous table above.
- The coefficients for the macro implementation variable show the effect of a one standard deviation increase

Table A-4. Correlation between macro implementation quality and macro design quality

Outcome is Macro implementation quality	(1) BW cred.	(2) FW cred.	(3) QF risks	(4) DSA	(5) Cons.	(6) Combined
Key variables						
Backward-looking credibility	0.263**					
	(0.103)					
Forward-looking credibility		0.526*				
		(0.274)				
Quasi-fiscal risks coverage			0.175**			
			(0.067)			
Debt sustainability coverage				0.180***		
				(0.055)		
Macro design consistency					0.032	
					(0.051)	
Combined macro design quality index						0.560***
						(0.093)
Project-level variables						
TTL previous projects	0.021	0.045	0.035	0.051	0.030	0.039
	(0.042)	(0.047)	(0.045)	(0.044)	(0.046)	(0.039)
DPO size	-0.059	-0.024	-0.025	-0.047	-0.042	-0.029
	(0.098)	(0.069)	(0.066)	(0.064)	(0.068)	(0.057)
Programmatic DPO	-0.107	0.026	0.086	0.132	0.121	0.095
	(0.132)	(0.125)	(0.111)	(0.108)	(0.112)	(0.097)
Econ. sector	0.183	0.231**	0.307***	0.216**	0.301***	0.110
	(0.123)	(0.114)	(0.103)	(0.103)	(0.102)	(0.105)
Country-level variables						
# previous DPOs	0.008	0.014	0.023**	0.014*	0.016*	0.017**
	(0.010)	(0.010)	(0.009)	(0.008)	(0.008)	(0.008)
IMF program overlap	-0.085	-0.232*	-0.258**	-0.314***	-0.288**	-0.233**
	(0.115)	(0.119)	(0.113)	(0.114)	(0.115)	(0.101)
GDP per capita (PPP)	0.202	0.062	0.050	0.147	0.129	0.090
	(0.222)	(0.165)	(0.125)	(0.122)	(0.128)	(0.116)
GDP (PPP)	0.075	0.065	0.007	0.083*	0.043	0.052
	(0.063)	(0.048)	(0.053)	(0.048)	(0.054)	(0.041)
Africa region	0.113	0.054	-0.113	0.038	-0.060	0.110

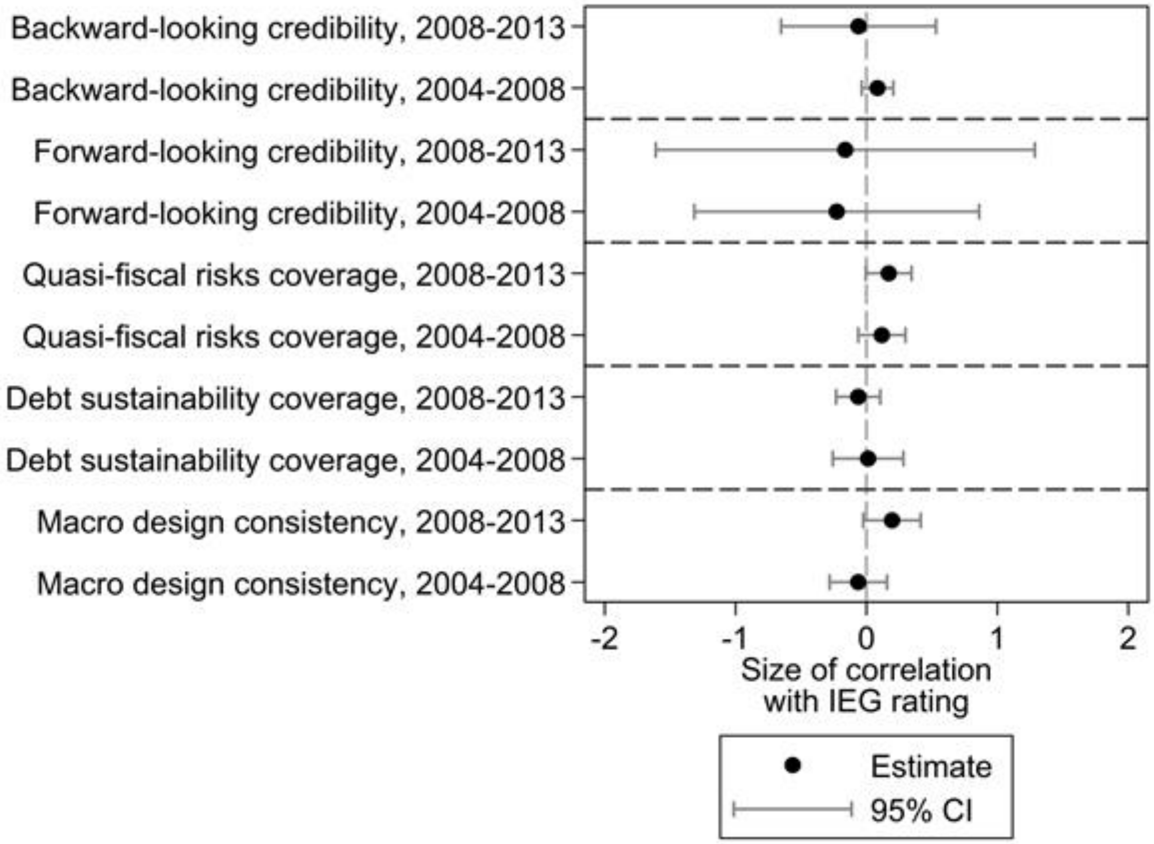
	(0.216)	(0.182)	(0.159)	(0.148)	(0.170)	(0.132)
CPIA score	-0.075	-0.102	-0.072	-0.284**	-0.230*	-0.112
	(0.130)	(0.145)	(0.131)	(0.122)	(0.124)	(0.109)
External shocks variables						
TOT change (avg. prev. 3 years)	-0.916	-1.275*	-1.142	-0.799	-0.641	-0.988
	(0.897)	(0.738)	(0.794)	(0.787)	(0.781)	(0.642)
GDP growth (avg. prev. 3 years)	0.007	0.011	0.007	0.019	0.009	0.018
	(0.021)	(0.022)	(0.019)	(0.019)	(0.019)	(0.016)
Constant	-2.399	-1.664	-0.220	-1.741	-0.818	-1.459
	(1.714)	(1.286)	(1.148)	(1.185)	(1.202)	(0.963)
Observations	123	153	173	174	174	174
Adjusted R-squared	0.075	0.085	0.159	0.139	0.090	0.235

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Outcome is implementation quality. Standard errors clustered by country-programmatic series.

This table shows the correlation between the macro implementation quality and the macro design quality.

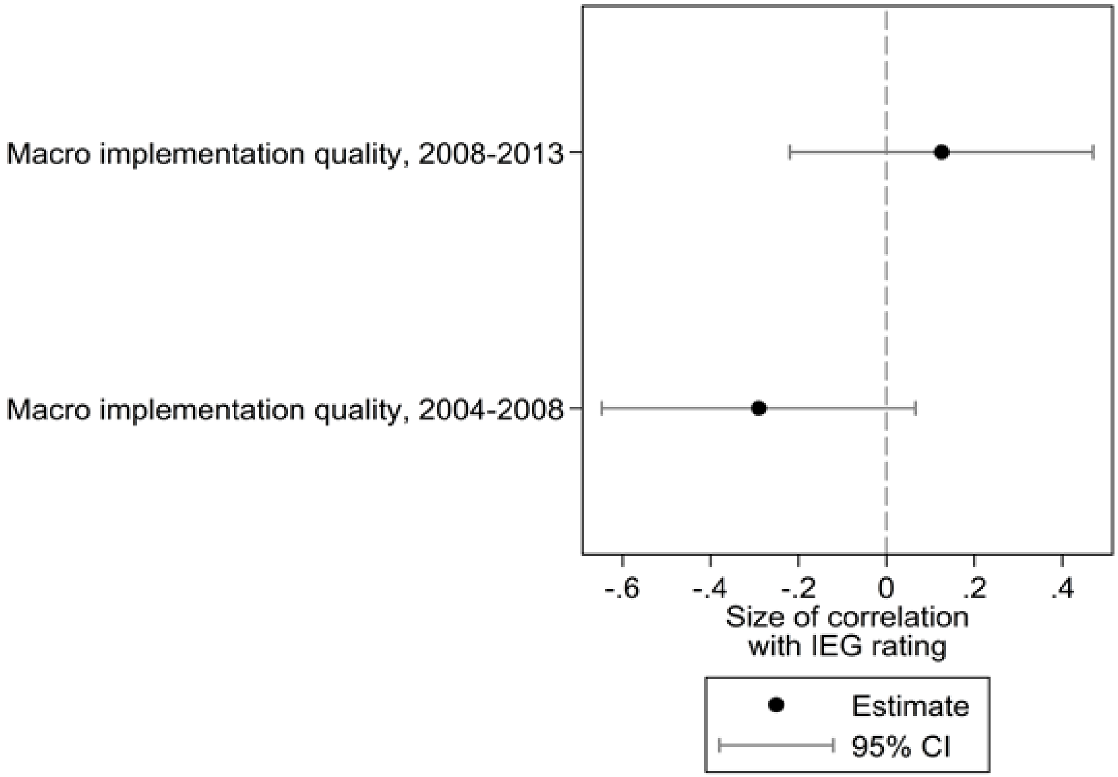
- The macro implementation quality is standardized to mean zero and standard deviation one. Each of the five variables is standardized to mean zero and standard deviation one. The combined index is a simple average of the five measures (therefore also having a mean of zero).
- First five columns show the correlation between five difference aspects of the macro design quality and the macro implementation quality, while controlling for other important factors. The last column shows the correlation between the combined index of these four measures and the macro implementation quality.
- The table shows that the quality of the macroframeworks design is generally positively correlated with the macro implementation quality (all but the macro design consistency measure).
- The coefficients for each of the variables shows the effect of a one standard deviation increase. For example, a one standard deviation increase in the backward-looking credibility is associated with a .26 standard deviation increase in macro implementation quality.

Figure A-3 Comparison of correlation of IEG ratings and macro design quality before and after 2008



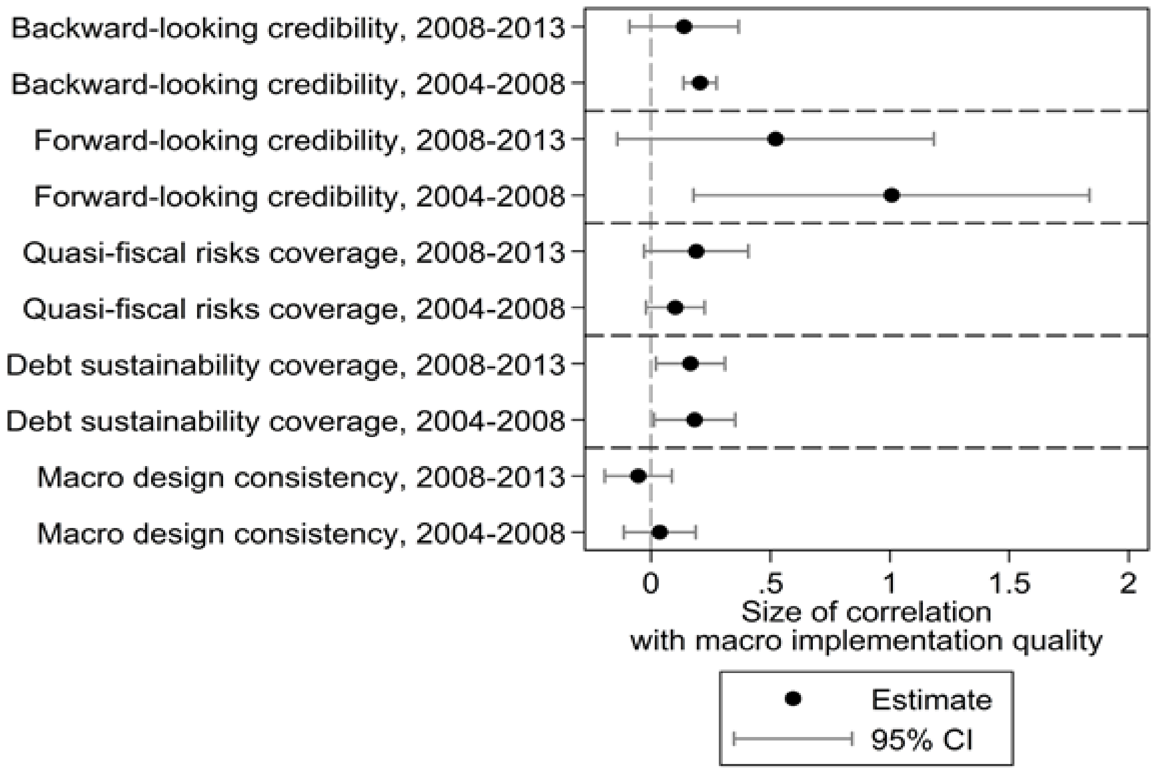
- This figure compares the size of the correlation between each aspect of the macro design quality and the IEG ratings before and after 2008, controlling for other important factors. Each dot in the graph is the estimate of the correlation, analogous to the estimates shown in the appropriate table above (i.e. these are simply the size of the coefficient from the regression). The horizontal bar around the point is the 95% confidence interval (that is, estimate +/- 1.96*standard error).
- Each panel, separated by the horizontal dashed line, shows the estimates after and before 2008 for the five different aspects of the macro design quality.
- The vertical dashed line denotes zero correlation.
- The graph shows that the correlation between the macro design quality and the IEG ratings does not show an apparent pattern across the two periods.

Figure A-4. Comparison of correlation of IEG ratings and macro implementation quality before and after 2008



- *The figure is analogous in logic to the one immediately before.*
  - The figure shows that the correlation between the macro implementation quality and the IEG ratings has somewhat increased since 2008, but that generally, it is not statistically different from zero (as the confidence interval crosses zero in both periods).

Figure A.5. Comparison of correlation of macro implementation quality and macro design quality before and after 2008



- This figure is analogous in logic to the previous two, but looks at the correlation between macro implementation quality and macro design quality.

There are no obvious changes in either direction across the two periods.

Variable explanations

Key variables

Backward-looking credibility: measures the absolute difference between the average actual fiscal balance (% of GDP) in years t-1, t-2, and t-3, and the planned fiscal balance in the first year of the program (year t). The actual fiscal balance is taken from the WDI database. The planned fiscal balance is taken from the WEO projections database. The projection refers to either the April or the October WEO projection in year t, whichever is closer to the project starting date. The final variable is standardized to mean zero and standard deviation one.

Forward-looking credibility: measures the average planned annual rate of change in the fiscal balance for years t, t+1, t+2, and t+3. The planned fiscal balance is taken from the WEO projections database. The projection refers to either the April or the October WEO projection in year t, whichever is closer to the project starting date. The final variable is standardized to mean zero and standard deviation one.

Macro design consistency index: originally ranges from 1 to 5, with 5 being highly satisfactory and 1 being highly unsatisfactory. The variable is rescaled to mean zero and standard deviation one.

Quasi-fiscal risk coverage: the measure is derived from a text search program. It represents the overall frequency count of the following acronyms and phrases in each project document: QFA, SOE, NPL, NPLs, quasi fiscal, contingent liability, fuel subsidy, cross subsidy, government guarantee, state owned enterprise, bad loan, non-performing loan, implicit debt, implicit pension debt, interest subsidy, fiscal cost of (bank) closure, and fiscal drain. The search was performed on text after removing punctuation, spaces, and after "stemming" the words, so that both singular and plural forms are returned. The variable is represented as the total frequency count of any of the mentioned phrases as the share of the total number of words in the document. The final variable is standardized to mean zero and standard deviation one. A small random inspection suggests high correlation of the measure with the quality of coverage of quasi-fiscal risks in the macro-fiscal frameworks.

Debt sustainability coverage: represents the overall frequency count of the following acronyms and phrases in each project document: DSA, debt sustainability, sustainability risk, debt sensitivity, shock to growth, and shock to debt. The search was performed on text after removing punctuation, spaces, and after "stemming" the words, so that both singular and plural forms are returned. The variable is represented as the total frequency count of any of the mentioned phrases as the share of the total number of words in the document. The final variable is standardized to mean zero and standard deviation one. Small random inspection suggests high correlation with the detail and coverage of the debt sustainability in macro-fiscal frameworks.

Combined macro design quality index: represents a simple average of the previous four measures.

Macro implementation quality: measures the average absolute difference between the actual fiscal balance (% of GDP) and the planned fiscal balance in years t, t+1, t+2, and t+3. The actual fiscal balance is taken from the WDI database. The planned fiscal balance is taken from the WEO projections database. The projection refers to either the April or the October WEO projection in year t, whichever is closer to the project starting date. The final variable is standardized to mean zero and standard deviation one.

Project-level variables

TTL previous experience: counts the number of previous DPOs for which the project TTL served as the TTTL. Drawn from the internal HR data.

DPO size: represents the logged dollar size of the DPO. Drawn from the DPO database.

Programmatic DPO: indicator variable taking the value of one if a DPO is programmatic and zero otherwise. Available only since 2005. Drawn from the DPO database.

Economic sector: indicator variable taking the value of one if a DPO sector board is Economic Policy.

Country-level variables

Number of previous DPOs: number of DPOs since 1990 a country had at start of a DPO. Drawn from the DPO database.

IMF program overlap: represents the share of the DPO duration that overlaps with an IMF program in the same country. Constructed based on the IMF data on country program dates and the DPO database.

GDP per capita (PPP): a country's GDP per capita in year t-1, expressed in constant 2011 dollars PPP. Drawn from the WDI database.

GDP (PPP): a country's GDP in year t-1, expressed in constant 2011 dollars PPP. Drawn from the WDI database.

Africa region: an indicator variable taking the value of one if the country is in Africa, and zero otherwise. Drawn from the DPO database.

Policy quality variables

CPIA score: CPIA score for the country at t-1, drawn from the confidential WB database.

External shocks variables

TOT change: average rate of change in the terms of trade index for years t-1, t-2, and t-3. Drawn from the WDI database.

GDP growth: average real GDP growth rate for years t-1, t-2, and t-3. Drawn from the WDI database.

Contemporaneous external shocks variables

TOT change: average rate of change in the terms of trade index for years t, t+1, t+2, and t+3. Drawn from the WDI database.

GDP growth: average real GDP growth rate for years t, t+1, t+2, and t+3. Drawn from the WDI database.

Oil price change: average change in crude Brent spot price for years t, t+1, t+2, and t+3. Drawn from the World Bank GEM database.