

2010:5

Sida Evaluation

Paolo de Renzio
Matt Andrews
Zac Mills

Evaluation of Donor Support to Public Financial Management (PFM) Reform in Developing Countries

Analytical study of quantitative cross-country evidence
Final Report





Evaluation of Donor Support to Public Financial Management (PFM) Reform in Developing Countries

Analytical study of quantitative cross-country evidence

Final Report

Paolo de Renzio

Matt Andrews

Zac Mills

Authors: Paolo de Renzio, Matt Andrews, Zac Mills.

The views and interpretations expressed in this report are the authors' and do not necessarily reflect those of the Swedish International Development Cooperation Agency, Sida.

Sida Evaluation 2010:5

Commissioned by Sida, Secretariat for Evaluation

Published by: Sida, 2010

Copyright: Sida and the authors

Date of final report: November 2010

Printed by: Edita, 2010

Art.no.: Sida61384en

ISBN 978-91-586-4167-9

ISSN 1401-0402

This publication can be downloaded/ordered from www.Sida.se/publications

Foreword

This study was prepared as part of a broader evaluation of donor support to public financial management (PFM) reforms. It analyses quantitative evidence on the quality of PFM systems, and assesses factors that may have determined cross-country differences and variations in the quality of PFM systems over time.

The cross country econometric analysis conducted as part of the study points to a number of findings. Countries with higher levels of per capita income, larger populations, and better economic growth are characterised by better quality PFM systems. In contrast, countries in conflict or post-conflict situations tend to have poorer quality of PFM systems. These economic and institutional factors explain most of the cross-country variation in PFM performance. The study also finds that, on average, countries that received more PFM-related technical assistance tend to have better PFM systems, but here the direction of causation could not be determined.

The study was constrained by a number of limitations and challenges, in particular poor data quality and limited information on donor PFM support. These limitations point to the need to interpret the results of the analysis with caution. The study also needs to be complemented with qualitative research at country level, which will be undertaken in the next phase of this evaluation through a number of country case studies.

The study was a collaborative effort between AfDB, Danida, DFID and Sida, and conducted by the Overseas Development Institute (ODI). On behalf of the management group¹ for the evaluation we would like to express our appreciation to the team leader Paolo de Renzio and the study team.

¹ The management group comprises: Joanne Asquith (AfDB); Ole Winckler Andersen (Danida); Nick Highton and Martin Aldcroft (DFID); and Katarina Kotoglou (Sida).

The findings and views are those of the study team and do not necessarily reflect the views of the AfDB, Danida, DFID, Sida, the management group, or its members.



Joakim Molander
Secretariat for Evaluation
Sida



Ole Winckler Andersen
Evaluation Department
Danida



Nick York
Evaluation Department
DFID



Colin Kirk
Evaluation Department
AfDB

Table of Contents

Acknowledgements	8
Acronyms and Abbreviations	9
Executive Summary	10
Introduction	13
Background, objectives of the study and analytical approach	14
Literature review and previous findings	17
Key variables and data collection	23
Data on PFM systems.....	23
Data on donor support to PFM reforms	31
Other variables.....	36
Results of the analysis: PEFA large-N sample	39
Results of the analysis: HIPC medium-N sample	52
Conclusions and implications for overall evaluation	55
References	60
Appendix 1: List of PEFA Countries	64
Appendix 2: List of HIPC/PEFA Countries	66
Appendix 3: PFM Clusters	67
Appendix 4: HIPC/PEFA indicators	68
Appendix 5: Donor data request form	70
Appendix 6: List of variables and summary statistics	72
Appendix 7: Terms of reference	75

Acknowledgements

The authors would like to thank the members of the Management Group, Ed Hedger, Greg Smith and Hafsa Mahtab at ODI, Joachim Wehner at LSE, participants at the World Bank seminar and donor officials who provided data and information.

Acronyms and Abbreviations

CPIA	Country Policy and Institutional Assessment
CRS	Creditor Reporting System
DAC	Development Assistant Committee
Danida	Danish International Development Agency
DFID	Department for International Development
GBS	General Budget Support
GDP	Gross Domestic Product
HIPC	Heavily Indebted Poor Countries
ICE	Imputation by Chained Equations
IDA	International Development Association
IFMIS	Integrated Financial Management Information System
IMF	International Monetary Fund
MTEF	Medium Term Expenditure Framework
ODA	Official Development Assistance
ODI	Overseas Development Institute
OECD	Organisation for Economic Cooperation and Development
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PEFA	Public Expenditure and Financial Accountability
PFM	Public Financial Management
PRSP	Poverty Reduction Strategy Paper
PSR	Public Sector Reform
Sida	Swedish International Development Agency
USAID	United States Agency for International Development
WB	World Bank
WLS	Weighted Least Squares

Executive Summary

This study is part of a broader evaluation of donor support to PFM reforms in developing countries. It brings together available quantitative evidence on the quality of PFM systems, to assess the factors that are associated with and may have determined cross-country differences and variations over time, with a particular focus on the impact of donor support for PFM reforms. The bulk of the analysis draws on data from PEFA assessments in 100 countries, data on donor support to PFM reforms directly collected from some of the donor agencies most active in this area, and a large dataset on other economic/social, political/institutional and aid-related variables that were identified as relevant from previous research.

There are a number of findings from the cross-country econometric analysis that are relevant for the evaluation, and for broader donor approaches and policies on PFM reforms. They can be summarised as follows:

- Economic factors are most important in explaining differences in the quality of PFM systems. Aid-related factors, on the other hand, have more limited explanatory power. As a consequence, PFM systems are more likely to improve responding to changing economic circumstances, rather than to donor efforts.
- More specifically, countries with higher levels of per capita income, with larger populations and with a better recent economic growth record are characterised by better quality PFM systems. On the other hand, state fragility, defined as being in a conflict or post-conflict situation, has a negative effect on the quality of PFM systems.
- Interestingly for the purposes of the evaluation, donor PFM support is also positively and significantly associated with the quality of PFM systems. On average, countries that received more PFM-related technical assistance have better PFM systems. However, the association is very weak: an additional 40–50 million US\$ per year would correspond to a half-point increase in the average PEFA score (equivalent to, say, a change from C to C+).

- These results remained consistent through a number of robustness checks and model changes. Interesting additional results come from using more recent data or focusing on low-income countries only. In these cases, the share of total aid provided as general budget support is also positively and significantly associated with better PFM quality. In other words aid modalities, and not just direct support to PFM reforms, contribute to explaining differences in the quality of PFM systems in some of the poorer countries where most donor efforts are concentrated.
- Finally, different aspects of donor support differ in their relationship with more specific PFM processes. A longer period of donor engagement, for example, is associated with better performance in upstream, *de jure* and concentrated processes. This may be due to donors' historical tendency to pay more attention to these simpler reform areas, but could also reflect the fact that downstream, *de facto* and deconcentrated processes take longer to improve.
- The level of donor PFM support is also more strongly associated with scores for *de jure* and concentrated PFM processes, again highlighting how donor PFM support seems to focus more on rules, procedures and specific actors within government. Results are reversed when it comes to upstream vs. downstream processes. Here, the association is stronger with downstream processes, possibly highlighting the large amounts of funding devoted to IFMIS projects, a typical downstream PFM reform.

At the same time, these results suffer from a number of serious limitations and challenges, including the following:

- Data quality remains an issue, especially when it comes to information about donor PFM support. Given the limitations of the information provided by donors, we focused on yearly disbursements for PFM-related activities. This gives undue weight to large projects such as IFMIS introduction, at the expense of 'softer' interventions. We also focused on data post-2002, for which availability is much greater. This means that we cannot capture earlier donor PFM support, when the foundations for PFM reforms were laid in some of the countries included in our sample.
- While the positive and significant relationship between donor PFM support (and GBS as a share of total aid in certain cases)

on one hand and quality of PFM systems on the other is particularly encouraging for the purposes of the evaluation, it clearly cannot be interpreted as causal given the nature of the data.

It could merely reflect the fact that donors tend to provide more PFM-related assistance (and more GBS) to countries that have already achieved a certain success in improving the quality of their PFM systems. Despite various attempts at tackling this issue, we could not prove the direction of causality.

- Assessing the impact of donor support on PFM reforms requires tracking the quality of PFM systems over time. Given the lack of sufficient time-series data, the analysis assumes that a higher PEFA score today is a valid proxy for past reform success. Its findings, however, are only partly confirmed by evidence from a smaller dataset looking at changes in PFM systems over the past decade in 19 African countries.

These limitations and challenges point to the need to interpret the results of the analysis presented in this paper with a lot of caution. Moreover, they highlight the need to complement these quantitative findings with in-depth qualitative research at country level, explaining not only *if* and *when* donor PFM support has had an impact on PFM systems, but also *why* and *how* it did. Case study countries, however, can and should be selected taking into account some of the insights provided in this paper.

Introduction

This Report is submitted by ODI to the Management Group for the “Analytical Study of Quantitative Cross-Country Evidence” in accordance with Contract No. C97027 signed with the Swedish International Development Cooperation Agency (Sida).

The work was carried out over the period from March to September 2010 by a team formed by Paolo de Renzio (University of Oxford and Overseas Development Institute, team leader), Zac Mills (independent consultant) and Matt Andrews (Harvard University). Ed Hedger managed and coordinated the contract at ODI. Greg Smith provided research support during the early stages of the project.

The report comprises the following sections:

- a) Background, objectives of the study and analytical approach
- b) Literature review and previous findings
- c) Key variables and data collection
- d) Results of the analysis: PEFA large-N
- e) Results of the analysis: HIPC medium-N
- f) Conclusions and implications for overall evaluation

Background, objectives of the study and analytical approach

This analytical study of quantitative cross-country evidence on Public Financial Management (PFM) in developing countries is part of a broader joint evaluation, initiated by the evaluation departments of Danida, Sida, DFID and the African Development Bank, in consultation with the OECD-DAC Evaluation Network. The evaluation aims to answer two sets of questions:

- a) Where and why do PFM reform efforts succeed?
- b) Where and how does external support to PFM reform efforts contribute most effectively to their success?

More specifically, the present study brings together all available quantitative evidence on the quality of PFM systems across countries and over time, to assess the factors that are associated with and may have determined cross-country differences and variations over time. One of the key issues for the study relates to how best to use available methodological tools to analyse existing quantitative data, in order to provide at least tentative or partial answers to the two evaluation questions. On one hand, the analysis aims to assess what country characteristics are associated with, or can be considered as causes of, successful PFM reform efforts. At the same time, it needs to focus more specifically on the contribution of donor support to PFM reforms to such success. Therefore, for question (a) the focus needs to be on contextual factors and conditions (economic, political, etc.) that can be used as independent variables, in order to assess their association/correlation with variations in the dependent variable (the quality of PFM systems) across countries and over time. For question (b), on the other hand, a somewhat narrower approach is needed, aimed at isolating the more specific impact of donor support to PFM reforms, while holding other factors constant.

In order to address these issues, while taking into account the scarce availability of data, the study focuses on two types of analysis. The first type of analysis (henceforth **PEFA large-N**) applies econometric techniques using the full cross-country dataset from PEFA (Public Expenditure and Financial Accountability) assessments

as the dependent variable, covering 100 countries (see Appendix 1). This large-N analysis takes advantage of the wider coverage and availability of cross-sectional PEFA data. It focuses on cross-country variation, and is based on two assumptions that are quite strong, and that need to be kept in mind when interpreting the results. First, that the PEFA methodology is a good way to assess and measure the quality of PFM systems. While PEFA indicators have been designed to assess countries' compliance with "good international practices" (PEFA 2005:5), there is evidence that "good public financial management means different things in different countries" (Andrews 2009:7). Second, that a higher PEFA score today is a valid proxy for past reform success. Given the lack of sufficient time-series data to assess changes over time, cross-country variation is assumed to depend mostly on the outcomes of recent PFM reform efforts, rather than on better initial conditions or other contextual factors. Multivariate regression techniques, designed to best fit the nature of the variables and the underlying model, are utilised to test the impact and significance of various factors on the quality of PFM systems across countries. For most of the explanatory factors, data refers to 2002–2006, to take into account possible lags in their effects and address some of the limitations of cross-country data.

The second type of analysis (henceforth **HIPC medium-N**) relies on the much more limited data that tracks changes in the quality of PFM systems over time. A smaller dataset was collated using HIPC1 and PEFA assessments for 19 countries in Sub-Saharan Africa (see Appendix 2)², including countries where repeat PEFA assessments have been carried out, in order to build a panel dataset that covers the period from 2001 to the most recent PEFA assessment. As such a medium-N sample is too small for any econometric treatment, simpler analytical methods are utilised, highlighting country clusters and relevant and interesting patterns or configurations of factors that are associated with PFM reform success, verifying and validating the findings from the PEFA large-N analysis.

-
- 1 The Highly Indebted Poor Countries (HIPC) initiative promoted a series of Assessments and Action Plans (AAPs) meant to track changes in the quality of PFM systems and identify necessary reforms for their improvement. For further details, see IDA/IMF (2003).
 - 2 For details of the method used to compile the dataset, see de Renzio and Dorotinsky (2007).

The combination of these two different kinds of analysis is the best possible way of using available data to shed some light on the evaluation questions in ways that go beyond the country-specific (but much more detailed) evidence that can be generated through the case studies. At the same time, it should be stressed that the findings from this study are limited to broad correlations and identification of relevant patterns, rather than strong and specific causal linkages. Nevertheless, we believe that the study makes an important contribution that does not only advance quantitative work on this subject, but can also nicely complement more in-depth qualitative case studies.

Literature review and previous findings³

In recent years, donor-supported PFM reform programmes have covered a range of initiatives aimed at strengthening the rules and procedures which underpin the budget process in aid recipient countries. These have typically focused on a number of standard interventions, which include improving the comprehensiveness of budget operations, building better links between annual allocations, medium-term policy objectives and performance indicators, and computerising budget management and expenditure controls⁴. What has certainly changed over time, however, is the scale of resources invested and the number of actors involved. A recent World Bank evaluation of Public Sector Reform (PSR) programmes (World Bank 2008), which include support to budget reforms, shows that the number of World Bank-financed projects with a substantive PSR component quadrupled between the early 1990s and 2005, increasing from less than 10% to more than 20% of total projects⁵. Data from the OECD DAC's database including all donors shows an even starker increase in committed funds for activities related to public sector financial management, which grew more than ten-fold, from US\$85.1m in 1995 to US\$930.6m in 2007. During the same period of time, the number of donor agencies involved in providing technical assistance in the PFM area has risen to over 25 (IMF 2007:22).

Given such interest and investment, it is somewhat puzzling that so little evidence and analysis exists on the comparative performance of PFM systems across countries and over time, on the factors that underpin successful PFM reforms, and on the role that donor agencies can play in PFM reform processes in developing countries. One of the key reasons for this, inevitably, is lack of available comparative data. As will be shown below, efforts at assessing the quality of PFM systems using standardized methodologies only started about a decade ago, when HIPC assessments were launched. In recent years, however, the introduction and gradual expansion

3 Parts of this section draw on a recent review in Hedger and de Renzio (2010).

4 World Bank (1998) and IMF (2007).

5 For Sub Saharan Africa, such proportion reaches 37%.

in the use of the PEFA methodology has provided researchers with more and better data to start addressing the evaluation questions.

There have been only two cross-country comparative analyses of PEFA assessment data so far, looking at the performance of PFM systems both across different areas of budget management and across countries. General analysis by de Renzio (2009a) of 57 PEFA assessments highlights how average indicator scores tend to deteriorate the further one moves through the budget cycle (from formulation to execution, reporting and scrutiny). Drawing on a dataset of disaggregated PEFA scores for 31 African countries, Andrews (2010) also investigates patterns or ‘themes’ in performance across PFM process areas. He reorganises the 73 PEFA indicator dimensions into clusters against the budget cycle. His first finding is consistent with that of de Renzio (2009a) for a wider span of countries. Average PEFA scores decline in the progression from upstream budget formulation to downstream financial management and accountability processes. On average, formal budget preparation and legislative budget review score most strongly, with external audit and legislative audit analysis shown to be among the weakest processes. The implication is that budgets are ‘better made than they are executed’.

Other interesting findings come from further distinctions that Andrews (2010) makes among PEFA indicators. First, he distinguishes PFM reforms linked to legislation, processes and procedures (i.e. *de jure* reforms), from those linked to the implementation or establishment of new practices (i.e. *de facto* reforms), finding that average scores for *de jure* dimensions are consistently higher than for *de facto* ones. In other words, improvements in budget practices lag behind reforms in budget laws and processes. Second, Andrews contrasts the performance of PFM process areas involving small groups of ‘concentrated’ actors, with processes which engage broader sets of ‘de-concentrated’ actors. Out of the total 64 disaggregated ‘budget cycle’ dimensions (excluding therefore those linked to budget outcomes), 26 are limited to concentrated actors such as the Budget Department or Debt Management Unit, while the remaining 38 dimensions relate to actors such as line ministries or Parliament. The evidence shows that countries score higher against the first set of measures, suggesting that actor concentration is associated with better functioning of PFM systems.

As far as cross-country comparisons are concerned, statistical analysis carried out by de Renzio (2009a) highlights some interesting cross-country patterns. For example, countries that are richer and more democratic have better quality PFM systems, while countries in Sub-Saharan Africa and South Asia perform worse on average. However, these binary associations are not necessarily significant from a statistical point of view, as for each country the overall average PEFA score might be determined by a number of these factors and categories interacting contemporaneously. Analysed through multivariate regressions, in fact, the only variables associated with significant changes in PEFA scores are income level and aid dependency. Even these findings are ambiguous. It is not surprising that higher income levels are significantly associated with higher quality of PFM systems. But it is not clear that income level *per se* is the driver of better PFM performance, rather than other variables that are often highly correlated with income, such as education levels or the share of government revenues that accrues to taxes rather than rents. The positive association with aid dependency, apart from the very small coefficient, may in fact reflect a reverse causality where countries with better budget institutions receive more aid.

In his analysis of PFM performance across African countries, Andrews (2010) uses a slightly different set of explanatory variables, such as: a) level of income and income growth; b) degree of country stability or fragility; c) dependence on ‘rents’ as major revenue sources; d) length of uninterrupted reform periods; e) type of administrative heritage. By organising the 31 African countries included in his analysis into five separate PFM ‘performance leagues’ according to their average PEFA scores, Andrews investigates the influence of each contextual variable upon PFM system strength. His findings reveal the following trends:

- a) The *economic growth rate* has a stronger association with higher quality PFM than the absolute level of income. In fact, some low-income but relatively fast-growing African countries feature in the highest PFM performance league.
- b) Country *stability* appears conducive to PFM progress. Fragile states – identified using an IMF classification – dominate the lowest league of PFM performance. These display particular weaknesses in strategic budgeting, budget transparency, budget execution and internal control.

- c) *'Rentier' states* (i.e. those which accrue most revenue from external sources, including natural resources, trade taxes and donor funding) tend to have weaker PFM systems compared with 'fiscal states' (i.e. those which collect a majority of their revenues from domestic citizens).
- d) Countries with a *PRSP*⁶ for more than three years achieve higher PEFA scores in almost all PFM process areas. The existence of a PRSP is used as a proxy measure for broad reform commitment, as it may 'lock in' pro-developmental policy choices and reform programmes.
- e) The evidence on *administrative heritage* is ambiguous, except for the downstream external accountability dimension, where Francophone countries tend to score lower against the PEFA indicators when compared to Anglophone ones.

The two studies by Andrews and de Renzio are useful in shedding some initial light on the factors that might influence the quality of PFM systems across countries, but say little about the role and influence of donors and donor assistance. Therefore, we looked at the broader literature on aid effectiveness, in particular at cross-country studies looking at the impact of foreign aid on recipient country governance. Again, there are just a handful of papers that look specifically at the effect of aid dependence and aid fragmentation on measures of the strength of institutions. Using different possible measures of state capacity, from indicators of bureaucratic quality to taxation efforts, Knack (2001) and Brautigam and Knack (2004) find that higher levels of aid dependence are indeed associated with declines in the quality of governance and in tax revenues as a share of GDP. This is how they justify this finding:

"the way large amounts of aid are delivered can weaken institutions rather than build them. This can happen through the high transaction costs that accompany aid, the fragmentation that multiple donor projects and agendas promote, problems of "poaching," obstruction of opportunities to learn, and the impact of aid on the budget process. Less directly, but just as important, high levels of aid can create incentives that make it more difficult to overcome the collective action prob-

6 A Poverty Reduction Strategy Paper (PRSP) is a comprehensive development plan that many African countries formulated in order to gain access to debt relief after 1997.

lems involved in building a more capable and responsive state.” (Brautigam and Knack 2004: 260–1)

These findings are supported by evidence in Moss et al. (2006), who emphasise how heavy reliance on external funding sources can generate an ‘aid-institutions paradox’, whereby as aid increases, the recipient government’s accountability relationship towards its citizens is weakened. Further evidence and tests provided by Ear (2007) also confirm the negative impact of aid on governance, even if just the technical cooperation component of aid is considered. The impact of aid dependency on revenue generation is slightly more controversial. Cross-country analysis presented in Moss et al. (2006) and Remmer (2004) presents further evidence in support of the argument that “high levels of aid dependence have failed to create strong incentives for governments to marshal new resources or developmental aims: instead, aid has simultaneously fostered the growth of government spending and the reduction of revenue effort” (Remmer 2004: 88). On the other hand, some case study evidence points to the fact that under specific circumstances aid flows are associated with a growth in domestic revenues (Fagernas and Roberts 2004, Pack and Pack 1990). Finally, Knack and Rahman (2007) argue that it’s not just the level of aid dependence that causes institutional deterioration, but also the degree of fragmentation of donor interventions, determined by how many donors are present in a recipient country and by their shares of the total aid that the country receives.

Most of the literature mentioned above uses an aggregate measure of aid dependence as the main independent variable, without recognising that foreign aid comes in many different forms and guise, which might have differential impacts on institutions. The only two more specific findings relate to technical cooperation and to aid fragmentation. In other words, it is important to keep in mind that the ‘quality’ of aid might be as important as its ‘quantity’ in determining its impact on governance and institutions in recipient countries. What about the more specific component of aid that this study aims to look at? At present, there are no comprehensive studies looking at the impact of donor support for PFM reforms on the quality of PFM systems. The only existing comprehensive evaluation of this kind of support is included in the World Bank evaluation mentioned above (World Bank 2008, Wescott 2009). The evaluation

finds that “about two thirds of all countries that borrowed for financial management showed improvement in this area“, with public financial management being “the most consistent area of improvement in the case studies” when compared to other aspects of public sector reforms (World Bank 2008:xv). More specifically, using CPIA data as a yardstick for improvements in the quality of budget institutions, the evaluation finds that 64% of countries that received any support for PFM reform programmes saw their CPIA PFM indicator score increase, compared to 32% in countries that did not receive such support (Wescott 2009:147). On the other hand, the evaluation also notes a number of problems with the way in which donors provided support to PFM reforms. For example, it notes how heavy donor involvement often leads to a situation in which “expectations and objectives [of budget reforms] tend to be more ambitious and global, reflecting the donors’ list of things that need fixing rather than the government’s list of things it is ready to do“ (World Bank 2008:40). The insistence “on a full array of public reforms“, the evaluation observes, means that “[World Bank] staff often lack the time and resources to develop a fully tailored product. So the result is likely to be one size fits all, off the shelf” (2008:41).

In summary, the existing literature that looks at determinants of PFM quality across the developing world, at the impact of foreign aid on governance and at donor support to PFM reforms is quite scarce, but nevertheless provides useful background and some interesting elements for our analysis. It identifies a number of variables that need to be included in our models. In terms of general country characteristics, it provides preliminary evidence of the importance of, for example, levels of income and income growth, strength of democratic institutions, government revenue sources, political stability and administrative heritage in shaping the quality of PFM systems. When looking at the influence of foreign aid, it highlights the role not only of overall aid dependency levels, but also of aid fragmentation in affecting governance standards in aid recipient countries. Important statistical issues linked to omitted variables and the possibility of reverse causation are also mentioned as issues that need to be considered in further research. Our study aims to build on these initial findings taking advantage of larger datasets that are becoming available. In the next section we therefore turn to issues of definition and measurement.

Key variables and data collection

The dataset that was compiled to carry out the analysis includes three sets of variables: (a) data on PFM systems (dependent variable); (b) data on donor support to PFM reforms (main independent variable); and (c) data on other independent and control variables. These are described in detail below.

DATA ON PFM SYSTEMS

There are very limited sources of information and cross-country data which can be used reliably to assess and compare the quality of PFM systems across countries and over time. For the large-N, cross-country analysis, the two datasets that we considered were the World Bank's Country Policy and Institutional Assessment (CPIA) and the set of indicators developed by the PEFA Secretariat.

CPIA's indicator 13 is produced by the World Bank as part of an annual internal performance rating exercise, and measures the 'Quality of Budgetary and Financial Management' along three dimensions: a) comprehensive and credible budgeting linked to policy priorities; b) effective financial management systems; and c) timely and accurate accounting and fiscal reporting. The indicator ranks about 75 countries on a six-point scale (1–6). Despite its relevance and coverage, the CPIA indicator suffers from two main drawbacks: (i) its single numerical value provides very limited detail and information on PFM system performance; and (ii) there is anecdotal evidence of CPIA ratings being based on subjective judgement and affected by lending decisions, therefore introducing important measurement errors⁷.

We opted instead for the use of PEFA data. The PEFA Performance Measurement Framework for PFM (PEFA 2005) is the most comprehensive attempt thus far at constructing a framework to assess the quality of budget systems and institutions. It comprises 28 indicators which assess institutional arrangements at all stages of the budget cycle, together with cross-cutting dimensions and indi-

7 See Arndt (2008).

cators of budget credibility. It also includes three additional indicators on donor practices. The dataset we worked with included the results of national-level assessments for 107 countries and territories. Of these, 7 countries were subsequently excluded from our sample: Kosovo and the Palestinian Territories because they are territories and not states, limiting the availability of other relevant data; Tuvalu because an important amount of other necessary data were not available; Norway because it is a clear outlier being a high-income country; and Bangladesh, Gabon and Nicaragua because the PEFA assessments had too large a number of missing indicators. The overall sample therefore includes data for 100 countries⁸. Two thirds of the assessments were carried out between 2008 and 2009. Only 42 of the 100 assessment reports have been made publicly available, while 24 are still considered to be at draft stage.

In order to transform PEFA scores into the dependent variable to be used in our large-N analysis, we followed a series of steps. First, we only considered indicators PI-5 to PI-28, as indicators PI-1 to PI-4 cover PFM system outcomes and performance, and not the quality of PFM systems per se. Second, for multi-dimensional indicators we used sub-indicator/dimension scores rather than summary indicator scores in order to fully exploit the information contained in the PEFA scores. This also allowed us to avoid the downward bias introduced by the M1 scoring methodology, where summary indicators are based on the lowest scoring dimension, or ‘weakest link’. Third, we converted the letter scores included in PEFA reports into numerical scores, with higher scores denoting better performance (from A=4 to D=1). Fourth, and finally, we constructed our dependent variable in three different ways:

- a) As an overall simple average of the 64 numerical scores that include all sub-indicators/dimensions for indicators PI-5 to PI-28;
- b) As averages of numerical scores for sub-indicators/dimensions in each of six clusters of indicators grouped by phase of the budget cycle⁹. This generates six sub-indices that will be used separately as dependent variables;

8 A full list is included in Appendix 1, alongside the year in which the assessment was carried out and the status of the report.

9 These clusters are slightly different from the ones included in the PEFA methodology, and have been rearranged to increase their level of internal consistency. For further details, see Appendix 3, Andrews (2010:8) and Andrews (2007).

- c) As individual scores for each of the 64 sub-indicators/dimensions in indicators PI-5 to PI-28. This generates a panel-type dataset of 64 dimensions * 100 countries.

It could be argued that these un-weighted averages are too simplistic an indicator to allow for cross-country comparisons. The PEFA Secretariat has also warned about a number of issues regarding aggregating PEFA scores and comparing them across countries (PEFA 2009). In order to check that values for the variable we constructed did not suffer from substantial biases, we used two procedures, one statistical and one substantive. For the statistical one, we imputed values for missing observations and applied Principal Component Analysis (PCA) to both the overall and the cluster averages. PCA is a statistical technique designed to detect the underlying structure of a number of related variables, and reduce their number through the creation of a new variable (or variables) that reflect that structure¹⁰. In our case, it generated an alternative ‘summary’ PEFA score based on the information included in the 64 underlying dimensions. The substantive procedure was based on the creation of a number of more parsimonious indices taking only sub-indicators/dimensions that donor assistance tends to focus on more directly into account, such as those linked to MTEFs, budget classification, internal controls, etc. In both cases, the resulting variables were very highly correlated with the overall average scores (> 0.95). We believe that this provides sufficient evidence that the overall averages used in our analysis in fact capture relevant aspects of the quality of PFM systems, and do not suffer from major biases. Finally, we think that the sample size is large enough to reduce the risk of invalid comparisons.

10 As described by Vyas and Kumaranayake (2006:460), “PCA is a multivariate statistical technique used to reduce the number of variables in a data set into a smaller number of ‘dimensions’.” Rabe-Hesketh and Everitt (2006) add that “the basic idea of the method is to try to describe the variation of the variables in a set of multivariate data as parsimoniously as possible using a set of derived uncorrelated variables, each of which is a particular linear combination of those in the original data.” Imputation by Chained Equations (ICE) was utilised to generate the 103 missing observations.

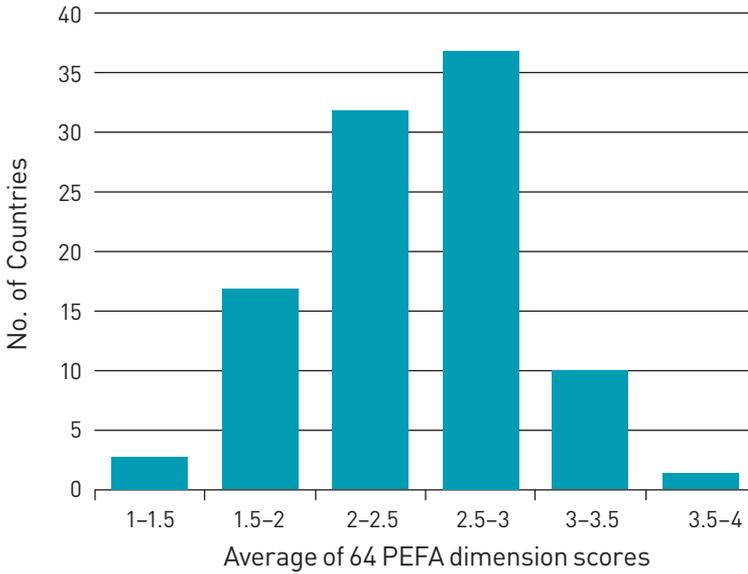
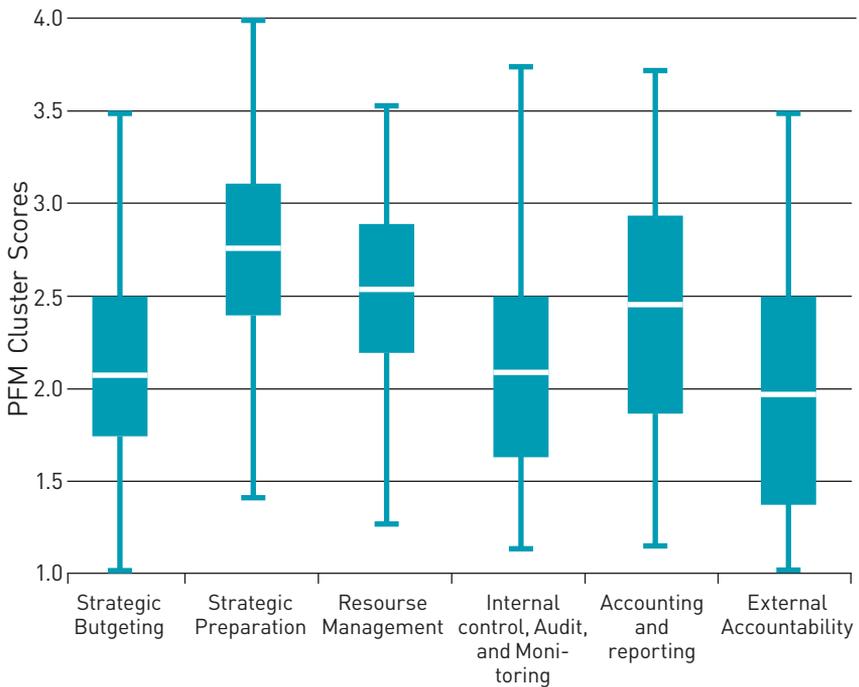


Figure 1. Distribution of overall average PEFA scores

Overall average PEFA scores across the 100-country sample vary between a minimum of 1.38 (Guinea-Bissau) to a maximum of 3.58 (South Africa), with a mean of 2.44 and a median of 2.47. Figure 1 shows the score distribution. As can be seen, most countries have average scores between 2 and 3, which in the original PEFA methodology would fall between a ‘C’ and a ‘B’ score. However, there are also 20 countries that would broadly score between a ‘D’ and a ‘C’, while there are only 11 countries with an average score of ‘B’ and above.

These aggregate averages inevitably hide a lot of the underlying information related to individual indicators and dimensions. To address this shortcoming, we have also looked at average scores across more specific PFM areas, created by grouping the 64 PEFA indicator dimensions in six internally consistent clusters that follow the budget cycle (see Appendix 3). These are: (a) strategic budgeting; (b) budget preparation; (c) resource management; (d) internal control, audit and monitoring; (e) accounting and reporting; and (f) external accountability (Andrews 2010:8). Figure 2 compares average scores and their distribution across these six clusters. As can be seen, ‘budget preparation’ is the area in which countries perform best, with an average score for the relevant PEFA dimensions of 2.77. The ‘resource management’ and ‘accounting and report-

ing' clusters are also characterised by results that are slightly above-average. Performance in the other three clusters is worse. In particular, 'external accountability' is characterised by the lowest average score for the relevant PEFA dimensions, 1.99. It has to be noted, however, that given the great variation within each cluster the differences between the average scores are not statistically significant.



Note: box-plots include information on the minimum, first quartile, median, third quartile and maximum average scores for the group of PEFA dimensions in each PFM cluster.

Figure 2. Average PEFA scores by PFM cluster

While these results confirm the previous finding that, for example, budgets are better made than executed, they also highlight two interesting exceptions. On one hand, the downstream 'accounting and reporting' cluster performs better than average, while the upstream 'strategic budgeting' cluster does not. This is particularly surprising, as 'strategic budgeting' refers mostly to the adoption of a medium-term perspective in budgeting, an area where donors have invested substantial resources. It may however reflect the

difficulties that countries face in introducing medium-term budgeting frameworks that goes beyond the mere projections of aggregate revenues and expenditures, but reflects well-developed sectoral plans and links capital spending with its recurrent implications.

We also looked at whether countries tend to score better or worse across the six clusters, by calculating the rank correlation coefficients for each pair of clusters. In other words, we wanted to check the likelihood that any country with a high average score in a cluster does better in other clusters, too. Surprisingly, coefficients are quite low, and range from 0.30 to 0.75. This indicates that countries do not necessarily score consistently across the various clusters. While we take this into account in the analysis that follows, this is certainly an area that deserves further attention and research.

Finally, we checked to see whether missing observations could be the source of any important biases. As indicated above, we excluded from the analysis three countries for which there were too many missing observations. While about half of the countries in our sample have some PEFA dimension that was not scored, in the great majority of cases this is limited to less than five of the 64 PEFA dimensions that we considered. Only two countries have more than 10 missing observations (Macedonia and Gambia). About half of the missing observations are due to the fact that PEFA assessments had a more limited scope and did not cover the full set of indicators, while less than one third are due to a lack of sufficient information to justify a specific rating. As far as specific indicators are concerned, the two indicators with the most missing observations (more than 15) are indicator PI-8 on transfers to sub-national governments and indicator PI-15 on tax collection. Otherwise, missing observations are quite evenly spread across the full range of indicators. Furthermore, countries with a higher number of missing observations are not significantly different from the rest of the sample in terms of their levels of income, region or any other relevant characteristic. In summary, we do not think that, after excluding countries with very incomplete assessments, missing observations constitute a serious measurement problem that might undermine the analysis that will follow.

For the second part of our analysis, focused on a medium-N sample of African HIPC countries, our aim was to address one of the main weaknesses of the PEFA dataset. While more than 150 assessments have been carried out since 2005, they provide only a snapshot of PFM system performance across countries¹¹. Interestingly, however, PEFA indicators can at least partially be mapped onto existing previous assessments, helping to extend their time series to more recent years. For the medium-N analysis, therefore, we combined PEFA data with data resulting from another methodology for assessing the quality of PFM systems, developed jointly by the IMF and the World Bank to test the systems in countries qualifying for debt relief under the Highly Indebted Poor Countries (HIPC) initiative. These HIPC Assessments were carried out in 2001 and 2004 in 23 countries¹², and scored country systems against benchmarks for fifteen indicators covering all stages of the budget cycle¹³. Despite some of the limitations of this methodology¹⁴, it constitutes the only codified ‘historical’ evidence that allows for a consistent tracking of the quality of PFM systems over time. On this basis, we compiled a small panel dataset that tracks changes in 11 indicators of PFM quality for 19 countries in Sub-Saharan Africa, many of which have also had two PEFA assessments (see Table 1 below). The dataset covers the period from 2001 to the most recent PEFA for each country, and groups indicators in three clusters, namely: (a) transparency and comprehensiveness (INFO), (b) linking budgets, policies and plans (POL), and (c) control, oversight and accountability (CTRL)¹⁵.

11 The total number includes repeat and sub-national assessments. As of October 2009, 21 countries had carried out repeat PEFAs, but often within a time span that is often not sufficient to see and justify substantive changes.

12 See IDA/IMF (2005). In 2006 three additional countries were assessed.

13 In 2004, an additional indicator on procurement was added.

14 First, calibration of the benchmarks did not capture the significant variations observed for some indicators across countries and over time. Second, the actual assessments in some cases revealed insufficient evidence to justify the scoring. Third, the indicators omitted important dimensions such as tax administration, fiscal decentralisation and parliamentary accountability.

15 For the details of which indicators fall under each cluster, see Appendix 4. For more detail on the methodology, see de Renzio and Dorotinsky (2007) and de Renzio (2009b).

Table 1. Quality of PFM systems across 19 African HIPC, 2001–2010

Country/ Year	INFO			POL			CTRL			TOTAL							
	2001	2004	2007	2010	2001	2004	2007	2010	2001	2004	2007	2010					
Benin	9	9	7	7	5	7	5	9	=	9	8	7	8	24	19	-	
Burkina Faso	8	8	10	11	7	8	7	7	-	8	9	9	10	+ 24	25	28	+
Cameroon	8	8	8	8	6	5	6	+ 8	+ 8	8	8	7	6	- 20	21	21	+
Chad	10	10	7	7	5	5	5	= 5	= 5	8	8	10	6	= 20	23	18	=
Ethiopia	8	8	8	8	7	5	7	+ 8	+ 8	8	8	10	5	+ 21	23	25	+
Gambia	8	8	8	8	4	5	4	- 8	- 8	5	5	5	5	- 21	17	17	-
Ghana	6	8	9	9	6	5	6	+ 4	+ 4	9	9	10	8	= 15	22	23	=
Guinea	8	9	8	8	4	4	5	+ 7	+ 7	8	8	6	6	= 19	21	19	=
Madagascar	10	10	10	9	5	6	5	= 7	= 7	7	7	7	7	= 22	23	22	=
Malawi	9	8	8	9	5	6	5	= 8	= 8	7	7	6	7	= 23	20	19	=
Mali	9	10	10	10	8	7	8	+ 10	+ 10	9	9	8	8	- 25	26	26	+
Mozambique	7	7	8	8	6	5	6	= 9	= 9	7	7	10	10	= 21	19	24	=
Niger	9	9	10	10	4	4	5	+ 6	+ 6	8	8	5	5	= 19	21	20	=
Rwanda	10	8	9	9	5	7	5	- 6	- 6	6	6	7	7	+ 23	21	21	-
S.Tome & Principe	8	9	8	7	3	3	3	+ 7	+ 7	6	6	4	5	= 18	18	15	=
Senegal	9	9	7	7	6	6	6	+ 8	+ 8	9	9	8	8	= 22	24	21	=
Tanzania	8	9	9	8	7	7	7	= 9	= 9	9	9	10	9	= 24	25	26	=
Uganda	10	8	8	8	7	6	7	= 8	= 8	9	9	8	8	= 25	23	23	-
Zambia	7	6	6	9	6	4	6	+ 7	+ 7	8	8	8	9	+ 17	18	20	+

Source: IDA/IMF (2005) and PEFA assessments. Based on authors' calculations.

Note: Numerical scores are based on methodology described in de Renzio and Dorotinsky (2007). '2007' denotes PEFA assessments carried out in 2005–07. '2010' denotes PEFA assessments carried out in 2008–10.

What the results show is that only five of the 19 countries (Burkina Faso, Cameroon, Ethiopia, Mali and Zambia) for which historical data is available saw an uncontroversial improvement in the quality of their PFM systems as measured by the subset of HIPC indicators. Four countries saw their PFM quality deteriorate (Benin, Gambia, Rwanda and Uganda). For the other ten countries, it is more difficult to detect a clear trend. Ghana, for example, recorded impressive improvements between 2001 and 2007, to then suffer a slight worsening of its overall score in 2010. Mozambique's overall score has improved over the whole period, but has seen some considerable fluctuations. Tanzania has consistently performed among the best in the 19-country sample, but has seen a recent slide. In terms of performance across the three different clusters, most improvements happened in linking budgets, policies and plans, where nine countries increased their score, while the area where the least progress was made was control, oversight and accountability.

The correlation with the overall PEFA averages used for the large-N sample is quite high (0.76), which ensures broad consistency between the two scoring methods. At the same time, there are some considerable differences. For example, the overall PEFA scores for Rwanda, Malawi and Mozambique are considerably higher using the 64 dimensions average than the sub-set of indicators that can be mapped back onto the previous HIPC assessments. These results therefore need to be taken with some caution, probably as describing broad (albeit somewhat incomplete) trends rather than specific changes in the quality of PFM systems.

DATA ON DONOR SUPPORT TO PFM REFORMS

Detailed and reliable data on donor support for PFM reforms is also difficult to find. The OECD-DAC Creditor Reporting System (CRS) database, the main global source of data on detailed aid flows, includes a sub-sector purpose code for 'public sector financial management' (15120). This includes information about donor commitments and disbursements for aid activities in support of PFM reforms, going back to 1995 for commitments, but only to 2002 for disbursements. Despite its apparent relevance as

a source for data on donor support to PFM reforms, the quality, reliability and comprehensiveness of the CRS data is highly questionable. Analysis of the underlying ‘micro-data’ (i.e. the specific entries submitted by donor agencies) reveals that not only a number of activities included should not be classified as support to PFM reforms, but also the omission of many activities that should be included under this classification. In other cases, the details available do not allow for the verification of the relevance of the aid activity for PFM reforms. This is especially true for multilateral agencies, such as the European Commission or the World Bank, which are not full DAC members and therefore have more limited reporting obligations.

These limitations constituted a serious challenge for the study. A substantive effort was therefore put into first-hand collection of donor data. To facilitate the task, we targeted a sub-set of 13 donor agencies that are particularly active in the field of PFM reforms. These included the members of the Management Group (Sida, Danida, DFID, and the African Development Bank), plus the Dutch, Norwegian and French aid agencies, the European Commission, the World Bank, the IMF, USAID, and the Asian and Inter-American Development Banks. While this sample does not ensure full coverage of data on donor support for PFM reforms, we believe that it provides a good picture and a suitable proxy for the amount of support provided to PFM reforms¹⁶.

Each agency was sent a data request form (Appendix 5) and asked to provide information about actual disbursements for technical assistance and other activities related to PFM reform support over the period 1995–2008. While limited information was requested for the 100 countries in the PEFA large-N sample, we asked for more detailed data for the medium-N sample of countries included in the HIPC/PEFA panel dataset (such as the PFM focus area, the main inputs provided, etc.). Of the 13 agencies contacted, ten replied, with varying degrees of completeness¹⁷. Unfortunately, no agency provided the more detailed information requested for the HIPC medium-N sample, which inevitably lim-

¹⁶ In fact, these donor agencies collectively provide more than 90 percent of total donor commitments for ‘Public Sector Financial Management’ as recorded in the DAC/CRS database for the period 1995–2008.

¹⁷ Data for Danida and French aid were in the end excluded as information was not sufficient.

ited the depth of our analysis. For the three missing agencies¹⁸, we looked at information available on public websites or went through the micro-data in their entries in the CRS database, selecting only the activities that could be identified as directly related to PFM reforms.

In a number of cases we had to make some judgement calls. Some agencies, for example (such as the World Bank, the African Development Bank and Danida), classify general budget support operations as PFM interventions. This greatly distorted the data, as we were looking for a measure of direct support to PFM reforms. Budget support, while it can indirectly strengthen PFM systems, is mainly aimed at financing general government operations. We therefore decided to either exclude these operations, or only include the share that was indicated as being directly associated with PFM reform initiatives, where this information was available. In other dubious cases, we simply trusted the data provided to us by donor agencies.

The yearly average of donor disbursements in each country was then included in our dataset as the main independent variable, based on the hypothesis that higher amounts of technical assistance should be associated with better quality PFM systems¹⁹. There are a number of problematic issues with this approach. For example, data might be driven by large projects such as the introduction of Integrated Financial Management Information Systems (IFMIS), which require substantial hardware investments, at the expense of ‘softer’ interventions that might affect PFM quality at a lower cost, as in the case of Medium Term Expenditure Frameworks (MTEFs). Given the greatly improved availability of disbursement data from 2002 onwards, we mostly utilised the sum of PFM-related disbursements over the period 2002–2006. While this makes sense in a number of ways, it also introduces a further potential source of bias in our data. On one hand, focusing on recent donor PFM support means that we are more likely to capture more direct effects on aspects of the PFM system that are part of the same consensus

¹⁸ European Commission, USAID and Asian Development Bank.

¹⁹ This goes against some of the more general findings in the literature about the negative effects of overall aid dependency, based on the assumption that technical assistance targeted at specific reforms does not suffer from the same drawbacks.

that is behind the construction of the PEFA methodology²⁰. On the other hand, our data do not include enough information on earlier donor PFM support, when the foundations for PFM reforms were laid in some of the countries included in our sample.



Figure 3. Total donor PFM support, 1995–2008

According to the data we collected, overall donor funding for PFM-related activities increased from around US\$60m in 1995 to more than US\$400m in 2008 (see Figure 3 above). The largest recipients of technical assistance in support of PFM reforms between 2002 and 2006 were Afghanistan (US\$33.2m per year) and Morocco (US\$25.1m per year). In Afghanistan’s case, the bulk of the funding came from the Asian Development Bank, the World Bank and DFID, funding a diverse set of PFM-related activities. In Morocco, PFM support is linked mostly to a series of World Bank loans for public administration reform, aimed at improving the effectiveness of public resource management. Other countries that received more than US\$15m per year over the same period are Indonesia, Tanzania and Mozambique. At the other end of the scale, the group of countries that received the smallest amounts (i.e. less than

20 Unfortunately, this might also generate a statistical problem of endogeneity. However, 2002–06 is the period preceding the time when most PEFA assessments were carried out, allowing us to at least partly address the issue of reverse causation (more on this below).

US\$10,000 per year) in donor support for PFM reforms over the same period include mostly small island states (St. Lucia, St. Kitts and Nevis, Kiribati and the Seychelles) and Belize²¹. The World Bank and DFID were the two donors, among those surveyed, that provided the largest total amounts in support of PFM reforms.

Interestingly, the correlation between our data on donor PFM support and data taken from the DAC/CRS database is a low 0.23. Given the effort that went into ensuring the reliability of the data we collected, we think that this is probably a reflection of the poorer quality of CRS data, something which should be taken into account in further analyses, and which highlights the need for donor agencies to increase the comprehensiveness and reliability of their CRS reporting.

Apart from direct support to PFM reforms, some additional variables were included in the dataset to capture other aspects of donors' impact on PFM systems in recipient countries. These are:

- a) A variable capturing the share of total aid being channelled through programme aid modalities (general budget support and sector programmes) calculated from DAC/CRS data, with the hypothesis that higher shares of programme aid are positively correlated with the quality of PFM systems;
- b) A variable looking at the length of donor engagement on PFM issues in each country, measured as the number of years passed since the first World Bank project in support of PFM reforms. The underlying hypothesis in this case is that longer engagement should be associated with better quality of PFM systems;
- c) Dummy variables for the presence of an IMF programme and a World Bank PRSC, in order to capture the presence of PFM-related conditionalities, hypothesising a positive impact of PFM-related conditionalities on the quality of PFM systems.
- d) An index of overall aid fragmentation calculated from DAC data, with the hypothesis that higher fragmentation impacts negatively on the quality of PFM systems²²;

21 It could be argued that per capita figures, or more generally figures adjusted for country size, provide a better indication of donor PFM investments. We nevertheless think that it makes more sense to focus on overall figures at this point, and control for country size in the analysis.

22 See Knack and Rahman (2007) both to see how the index is calculated and for the main arguments behind the hypothesis.

OTHER VARIABLES

Apart from the two main variables of interest – ‘quality of PFM systems’ (dependent variable) and ‘donor support for PFM reforms’ (independent variable) – a number of other independent variables were included in the analysis. They represent other factors which can be hypothesised as having an influence on the level and change of the dependent variable, including those which have also been identified as relevant factors by the previous empirical analyses summarised above. For ease of reference, these variables are summarised in Table 2 below, distinguishing between economic/social and political/institutional variables, and specifying the hypothesised impact on the quality of PFM systems.

Table 2. Other independent variables included in the PEFA large-N analysis

Variable	Effect ^a	Detailed hypothesis/References
Economic/social variables		
GDP per capita	+	Richer countries have better government capacity and a better-educated population which will hold government accountable for how it manages public finances. See de Renzio (2009a)
Recent GDP growth	+	Economic performance and increasing resources allow for increasing fiscal and reform space necessary for PFM reforms. See Andrews (2010)
Aid dependency	-	Aid dependency worsens governance standards and distorts incentives for reform. See Knack (2002) and Brautigam and Knack (2004)
Resource dependency	-	The ‘resource curse’ hypothesis claims that dependence on natural resource revenues undermines PFM and other governance reform prospects. See Andrews (2010) and de Renzio et al. (2009).
Population	?	There could be economies of scale in investing in PFM systems in larger countries; otherwise public finances might be easier to manage in smaller countries. See de Renzio (2009a)

Variable	Effect ^a	Detailed hypothesis/References
Adult literacy	+	In countries with a better educated population, government will have better capacity, and citizens will hold government accountable. See Kaufmann et al. (1999) ²³
Trade openness	+	Open economies need to be better managed, and are more exposed to external pressure for better economic management. See Treisman (2000) and Busse et al. (2007)
Technological diffusion	+	Electric power consumption, telephone subscriptions and use of personal computers and the internet all make it easier for governments to adopt PFM reforms and for citizens to hold government accountable. See Kakabadse et al. (2003)
Political/institutional variables		
Administrative heritage	?	Inherited budget systems affect present-day PFM quality. See Andrews (2010) and La Porta et al. (1999)
Democracy	+	In democratic regimes, citizens will hold government accountable and demand better management of public resources. See Wehner and de Renzio (2010) and Rivera-Batiz (2002)
State fragility/conflict	-	Countries in fragile or in conflict/post-conflict situations have more difficulties in carrying out PFM reforms. Capacity is very weak, informality is predominant, and political will is lacking. See Andrews (2010)
Regional dummies	?	Regional dummies capture other country characteristics, such as geography and culture, that may affect PFM reforms.

a) Expected sign of the regression coefficient. Question marks denote cases where this is not clear, or where the variable is categorical.

While measuring most of the variables included in the table above is not problematic, for ‘democracy’ and ‘state fragility’ measurement is less straightforward. A number of alternative measures of democ-

²³ Kaufmann et al. (1999) note the correlation between adult literacy and governance, but argue that the causal link goes from the latter to the former.

racy exist, and debate around their usefulness is very much alive²⁴. We settled for one of the most commonly utilised measures by Freedom House, but also checked whether our results changed when other indicators were used instead. For state fragility, we decided to use a dummy variable recording the presence of UN peacekeepers in the country since 1995, therefore equating state fragility with a conflict or post-conflict situation²⁵.

Further details on measures utilised and summary statistics for all variables are reported in Appendix 6. We were unable to include in our analysis a number of other potentially relevant variables, related for example to levels of corruption and government tax effort. This was due to the fact that data sources (e.g. the International Country Risk Group) did not cover a sufficient number of countries.

24 See, for example, Munck (2009) and Cheibub et al. (2010).

25 Other definitions of fragility, such as those used by the World Bank or in the 'State Fragility Index' produced by the Polity IV Project, in our view are too broad to capture the key elements of fragility we are interested in.

Results of the analysis: PEFA large-N sample

The average PEFA scores that were used as the variable to be explained can be considered to represent a measure of the overall quality of PFM systems. Assuming that the present-day quality of such systems accurately reflects the results of past reform efforts, what can be said about the factors that determined successful reform efforts across our sample of 100 countries? And more specifically, what was the impact of donor assistance in support of such efforts? The main objective of the PEFA large-N analysis was to provide some tentative answers to these questions. Most PEFA assessments from which PFM quality scores are derived took place in 2008–9 (see Appendix 1). For our explanatory variables, we used data that cover the five years between 2002 and 2006, in order to assess the impact of past conditions and activities on present PFM systems and due to some of the data limitations described in the previous section.

Given the mostly cross-sectional nature of our data (meaning comparing data across countries, and not over time), the standard econometric method to be used is Ordinary Least Squares (OLS) regression. In this section, we report on the analysis that was carried out and its results, including a number of robustness checks to verify their strength against a number of alternative models and measures, and highlighting their limitations and the caution needed in interpreting them.

As a first step, we looked at some bivariate correlations, to see the extent to which our data confirmed both previous findings from the literature and some of our preliminary hypotheses. Figure 4 below presents scatter-plots of average PEFA scores against income levels, income growth, aid dependency and donor PFM support over the period 2002–2006. Over our 100-country sample, data show positive correlations with income levels, income growth and PFM aid, and a negative correlation with overall aid dependency, as expected, despite large variance.

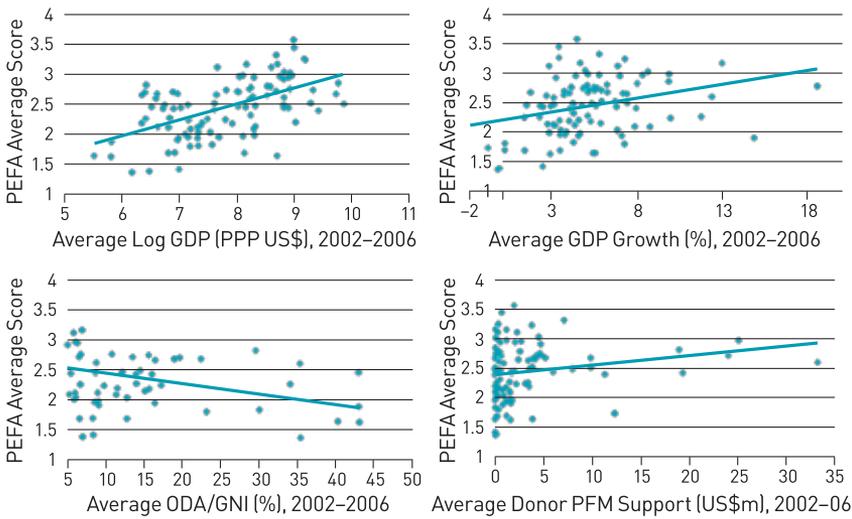


Figure 4. Average PEFA Scores and key explanatory factors (bivariate scatter-plots)

We then turned our attention to multivariate analysis, to understand how these and other variables identified jointly affected PFM quality. We first ran a series of exploratory regressions using sub-groups of variables to establish which economic, political and aid-related variables were significantly associated with changes in the quality of PFM systems and should be kept in subsequent stages of the analysis. The results are shown in Table 3.

Table 3. Preliminary regressions by groups of variables

Dependent Variable PEFA Avg. Score	Economic/ social variables	Political/ institu- tional vari- ables	Aid- related variables
GDP per capita (log)	0.32***		
Recent GDP growth	0.04***		
Population (log)	0.08**		
Resource depend- ency	-0.37***		
Adult literacy	0.001		
Trade openness	-0.002**		

Dependent Variable PEFA Avg. Score		Economic/ social variables	Political/ institu- tional vari- ables	Aid- related variables
Technological dif- fusion	Electricity Comput- ers	0.00 0.00		
Observations		94		
R-squared		0.57		
Admin. heritage	British colony French colony		-0.15 -0.26*	
Democracy			-0.04	
State fragility			-0.38***	
Regional dummies	SSA LAC ECA		-0.07 0.09 0.31**	
Observations		100		
R-squared		0.33		
Aid dependency				-0.02***
PFM support				0.24***
Aid fragmentation				0.17
GBS as % of total ODA				0.24
IMF programme				0.02
WB PRSC				0.14
Length of donor eng.				-0.003
Observations		97		
R-squared		0.25		

Notes: Ordinary Least Squares regressions. Only coefficients are reported.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

Among technological diffusion variables, we dropped telephone subscrip-
tions and internet users as they were highly correlated with the others.

Economic and social variables are the ones that explain a larger
part of the variation in the quality of PFM systems (shown by

a higher R-squared value of 0.57), and most of their coefficients have the expected sign. Higher levels of GDP per capita and of recent economic growth are positively and significantly associated with better PFM systems, as is the size of the country's population, whereas dependency on natural resource revenues and trade openness have negative coefficients. While this result confirms the standard 'resource curse' hypothesis, that links resource-dependent countries with worse governance standards, the sign for the trade openness variable is not in the expected direction. At this preliminary stage, however, it is not worth advancing further explanations, as results might be driven by omitted variables. Among political and institutional variables, the ones with a significant coefficient (apart from the Eastern European regional dummy) are the dummies for former French colonies and for conflict-affected states, both with negative coefficients that confirm the findings of previous research. The level of democracy also comes close to standard significance levels, also with the expected sign²⁶. Aid-related variables, in this preliminary phase of the analysis, only explain a quarter of the differences in the quality of PFM systems (R-squared=0.25). Results highlight the negative and significant association between the overall level of aid dependency and PFM quality and, interestingly, a positive and significant one of donor PFM support. Another variable that comes very close to standard levels of statistical significance, namely the presence of a PRSC, also has the expected positive coefficient.

In the following step of the analysis, we built the main model to be used for our analysis by bringing together the variables that were shown to be significant in the exploratory regressions, plus a few others of specific importance, such as some of the key aid variables, regional dummies (to control for various regional characteristics), and the level of democracy, covering the period from 2002 to 2006. The results are shown in Table 4, first using standard OLS regression, but also using Weighted Least Squares (WLS) to address issues of heteroskedasticity, which persisted even when we used the robust standard errors option.

Some of the key results from the preliminary regressions also hold in the comprehensive model. Once again, higher levels of income,

²⁶ The negative sign is due to the fact that Freedom House classifies stronger democracies with lower values of the variable.

income growth and population are significantly associated with better quality PFM systems. Countries that are richer, larger, and have a good economic growth record in recent years are also more successful at reforming and improving their PFM systems. More specifically, countries with double the income per capita can be expected to have an average PEFA score which is almost half a point higher, holding other factors constant. Similarly, countries which are poorer and with a worse recent performance in terms of economic growth, will also be characterised by PFM systems of lower quality. Donor support for PFM reforms retains a positive (although slightly less significant) coefficient. Its low value, however, indicates that very large injections of technical assistance would be coupled with only a small increase in PEFA average scores. According to the analysis, an additional US\$40–50m per year in PFM assistance (that is, more than doubling the amount received by the country that has received the most in recent years, Afghanistan) would be associated with just a half-point change in the overall PEFA average score, holding other factors constant. Finally, the presence of peacekeepers as a proxy for state fragility maintains its significant negative association with PFM quality, highlighting the importance of political stability for PFM improvements.

Table 4. The determinants of the quality of PFM systems

Dependent Variable PEFA Avg. Score	OLS	WLS
GDP per capita (log)	0.44***	0.45***
Recent GDP growth	0.02**	0.02*
Population (log)	0.08**	0.08***
Resource dependency	-0.19	-0.19
Trade openness	-0.001	-0.001
French colony	-0.09	-0.09
Democracy	-0.02	-0.01
State fragility	-0.22**	-0.21*
Aid dependency	-0.01	0.01
PFM support	0.01**	0.01*
GBS as % of total ODA	0.76	0.71
Length of donor engagement	0.01	0.01
SSA dummy	0.11	0.10

Dependent Variable PEFA Avg. Score	OLS	WLS
LAC dummy	0.02	0.009
ECA dummy	0.17	0.15
Observations	93	93
R-squared	0.68	0.68

Notes: Only coefficients are reported.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

These initial results shed some light on the factors that drive differences in cross-country PFM performance. For the purposes of the overall evaluation, the positive and significant relationship with PFM-related technical assistance is particularly encouraging, though it clearly cannot be interpreted as a causal. It could merely reflect the fact that donors tend to provide more PFM-related assistance to countries that have already achieved a certain success in improving the quality of their PFM systems. The issue of endogeneity also needs to be taken into account in interpreting the relationship between income levels and PFM quality. We come back to this problem further below.

In order to better understand the nature of the possible impact of donor PFM support on average PEFA scores, we introduced some interaction terms in the basic model. These were aimed at testing whether donor PFM support was in some way linked to the presence of higher levels of programme aid or of conditionalities linked to IMF or World Bank programmes. None of these interaction terms, however, reached conventional significance levels.

We then used the same model to explain variation not in the overall PEFA average score, but in each one of the six PFM cluster scores linked to the different phases of the budget cycle. While there inevitably are differences in the results (some of which are more difficult to interpret), the overall trends did not change dramatically. Some results are worth highlighting: (a) donor PFM support retains a positive and significant association with average PEFA scores in most of the six PFM clusters; (b) coefficients for income levels and growth maintain their positive sign and significance in most clusters, except for 'strategic budgeting'; (c) in 'strategic budgeting', GBS as a share of total aid becomes significant, possibly indicating that successful medium-term budget projections

are more successful in those countries where donors channel more of their aid through government channels²⁷; (d) ‘external accountability’ scores in former French colonies and in Latin American countries (which are all former Spanish or Portuguese colonies) are significantly lower²⁸.

A series of robustness checks were then carried out in order to address a number of issues related to the data and the nature of our variables, and to check for consistency with the results we obtained in our main model.

First, we looked at the possibility of measurement errors biasing our results by repeating the analysis using alternative measures for a number of our variables, such as democracy, state fragility, and the use of aid modalities. For democracy, we used both Freedom House and PolityIV as sources of data to assess whether the strength of democratic institutions had any effect on PFM quality. Similarly, we looked at different existing indices of state fragility, and expanded our definition of programme aid to include both general budget support and sector programmes. In all cases, our results were not affected. Interestingly, however, when we substituted the PFM support variable that we constructed from donor data with similar data from the DAC/CRS database, the statistical significance of the coefficient vanished. This is not surprising, given the low correlation between the two variables reported above. We interpret this as a further confirmation of the worse quality of the information contained in the CRS database.

Second, we ran an identical set of regressions using data for our explanatory variables averaged over the period 2004–08 rather than 2002–06, to test whether our results were dependent on the time period considered. Again, the overall story holds, with the additional finding that over this more recent period the share of GBS in total aid flows is also significantly and positively associated with the qual-

27 It should also be noted, however, that the R-squared statistic for the ‘strategic budgeting’ regression is the lowest of the six clusters (0.29), indicating that other variables might better explain variations in this PFM cluster.

28 This result could be interpreted in two different ways: either external audit is really less effective in countries characterised by a civil law (as opposed to a common law) legal background, or the PEFA methodology is not designed to adequately assess the external audit systems that are prevalent in civil law countries, having been conceived with a common law, Westminster-style government in mind.

ity of PFM systems. Rather than provide evidence in support of the positive impact of budget support, however, this might more simply reflect the increasing importance of GBS as an aid modality in recent years, and the fact that donor agencies reward countries with better PFM systems by shifting more of their aid to directly support the government budget.

Third, the conversion of PEFA scores from letter to numerical scores and their subsequent averaging means that we treated a variable that is ordinal in nature as if it were a continuous variable. Using OLS with an ordinal dependent variable is known to lead to biased results. One of the statistical techniques that correct for this problem is ordered logit regression. For this purpose, we used individual PEFA dimension scores, rather than averages, as our dependent variable (as in Andrews 2010). This allowed to also assess whether countries scored higher on *de jure* rather than *de facto*, upstream rather than downstream, and concentrated rather than deconcentrated PFM processes, testing the findings reported in Andrews (2010) for a larger sample of countries. These differences prove to be significant, confirming the results of past research that countries make budgets better than they execute them, pass laws better than they implement them, and progress further when responsibility for reforms lies within core groups in the Ministry of Finance. Moreover, even after controlling for these differences, the results linked to our other variables still hold. The positive and significant coefficients for income levels and income growth remain, pushing countries up a notch in their ordinal PEFA dimension scores. Higher donor PFM support is associated with higher average PEFA scores in this case, too, while the GBS share of total aid seems to be associated with positive changes in PFM quality for countries with lower PEFA dimension scores.

Fourth, we wanted to check that our results were robust to sample changes, and not driven by specific outliers or specific groups of countries across the various variables included. We therefore repeated our analysis with the following changes: (a) excluding very small countries (with populations of less than 1m), as these often skew results; (b) excluding, in turn, middle-income countries and low-income countries; (c) excluding countries with particularly high values of donor PFM support (in excess of US\$15m per year over the period 2002–06); (d) excluding countries where PEFA assessments

were still in draft form, to ensure that the lower quality of such assessments was not unduly affecting results. Results for the main variables of interest are reported in Table 5 below.

**Table 5. Partial regression results
(robustness checks with sample changes)**

Dependent Variable PEFA Avg. Score	Excl. small countries	Excl. MICs	Excl. LICs	Excl. PFM outliers	Excl. draft PEFAs
GDP per capita (log)	0.42***	0.35***	0.50***	0.43***	0.33***
Recent GDP growth	0.03**	0.02*	0.03**	0.02**	0.02
Population (log)	0.03	0.07	0.03	0.08**	0.06*
State fragility	-0.24**	-0.27**	-0.22	-0.23**	-0.21*
PFM support	0.01**	0.01**	0.01	0.01	0.01*
GBS as % of total ODA	0.71	1.21*	0.71	1.03*	0.94
Observations	75	70	58	89	71
R-squared	0.73	0.59	0.66	0.68	0.73

Notes: Ordinary Least Squares regressions with robust standard errors. Only coefficients are reported.

All regressions include all other variables from main model in Table 4.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

Once more, results showed to be reasonably robust to many of these changes. Most of the variables that were found to be significantly associated with changes in the quality of PFM systems maintain the direction and the significance of their coefficient. Some results deserve further comments, however. First, the positive association between PFM quality and donor PFM support seems to be driven by the small number of countries that receive large amounts of PFM assistance. When the five countries receiving more than US\$15m per year (Afghanistan, Morocco, Indonesia, Tanzania and Mozambique) are excluded from the analysis, the PFM support variable loses significance. Second, when only poorer countries are considered, income levels become less important in explaining differences in PFM quality, while increasing shares of GBS in total aid become positively and significantly associated with better PFM quality. Given that the analysis applies to a more homogeneous group of countries, the result for income

levels can be expected. The result for GBS is more interesting, as it indicates that in low-income countries it is not just PFM support, among donor interventions, that is significantly associated with changes in the quality of PFM systems. Aid modalities also count. Once again, however, endogeneity issues limit the conclusions that can be drawn from this interesting finding. Third, state fragility and PFM support are significant factors among poorer countries. When low-income countries are excluded from the analysis, both variables lose significance.

Finally, we needed to address the very important issue of reverse causation for some of the key explanatory variables that were found to be significant. As indicated above, the relationship between both income levels and donor PFM assistance on one side, and the quality of PFM systems on the other (which is significant, positive and robust to a number of specification changes) cannot be interpreted as causal. While it could be that higher levels of income and donor PFM support bring about improvements in PFM systems, the opposite could be just as true. In other words, better quality PFM systems might explain, rather than be explained by, higher GDP per capita and higher levels of donor support for PFM reforms. More specifically for PFM-related aid, what our results might be capturing is the tendency of donor agencies to invest more money in countries that have already shown the capacity and willingness to reform their PFM systems, and that have gone past simpler initial reforms and are starting to engage with more complex (and more expensive) reforms.

Such endogeneity issues are very difficult to address with the kind of data that we had at our disposal, which have no time-series dimension that can shed light on how PFM systems have evolved within each country. Nevertheless, we looked into the possible use of the statistical technique most used to address reverse causation, which relies on the identification of so-called ‘instrumental variables’. These are variables which are highly correlated with the endogenous explanatory variable but do not suffer from the same problems with relation to reverse causation. In our case, the challenge was to find adequate instrumental variables for both income levels and donor PFM support.

In the case of income, we found that 1995 levels of GDP per capita and infant mortality explain most of the differences in the more

recent income levels used in our analysis, without being correlated with the quality of PFM systems in 2008. Using these two variables as instruments, we find that a two-stage least squares procedure generates results that are very similar to the OLS ones, which means that we can more comfortably claim that it is higher income levels that bring about better PFM quality, and not the other way around.

When it came to disentangling the endogeneity of donor PFM support, unfortunately we could not find any suitable instrumental variable that could help us determine in which direction the causation really works. For example, we tried using: (a) administrative heritage dummies, as they may reflect not only general donor interest, but more specific interest in supporting institutional reforms (as many institutions have colonial origins); (b) indicators of the past quality of institutions, as this could have shaped donors' intentions to support institutional development without being highly correlated with PFM quality today; (c) the degree of economic and political stability over the past 10–15 years; (d) dummy variables for the existence of a PRSP or for HIPC status, as much of the original decision to allow countries to enter HIPC were linked more to macro-stability and indebtedness rather than to PFM quality; (e) other variables linked to the length and depth of donor engagement with broader public sector governance. None of these variables, however, could explain, individually or jointly, recent levels of donor PFM support. While this may be a reflection of the complex inter-relationships that exist in reality between donor and government efforts at reforming country institutions, it may also be due to the biases we noted above in the data that were collected for the donor PFM support variable. The nature of donor PFM interventions, moreover, does not lend itself easily to statistical analysis that treats all countries as equal cases. Country idiosyncrasies might play an important role in explaining the difficulties we faced in interpreting the results of the analysis. In some cases, for example, large donor PFM support might be due to an expensive computerized financial management system that is at all adequate to the context, and therefore fails to generate improvements in the overall quality of PFM systems. In other cases, smaller interventions may be much more contextually relevant, leading to more important improvements.

A final set of regressions focused more specifically on how different aspects of donor engagement in PFM reform processes affected different PFM process types, as defined by Andrews (2010). In other words, we wanted to check if there were any significant linkages between variables such as donor PFM support and share of programme aid, and certain aspects of PFM quality, i.e. those more directly related to upstream rather than downstream processes, to *de jure* rather than *de facto* reforms, and to PFM processes characterised by actor concentration rather than deconcentrated responsibility. We then calculated PEFA average scores for the indicator dimensions belonging to each of the types of PFM processes, and used them as dependent variables. Table 6 reports the results.

While GBS as a share of total aid does not bear any direct relationship with any of these specific types of PFM processes, some interesting patterns can be seen for donor PFM support and the length of donor engagement. First, a longer period of donor engagement is significantly and positively associated with upstream, *de jure* and concentrated processes. These results may be related to the fact that donors have historically paid more attention to these simpler reform areas, rather than getting involved in the messier, and more difficult, challenges of implementing PFM reforms that address problems in government-wide budget execution. This, in turn, could explain the very small impact of PFM support on overall PFM quality. They could also stem from the fact that downstream, *de facto* and deconcentrated processes take longer to develop and improve, with donor engagement so far not being sufficiently long and sustained to be associated with substantive improvements. Second, the level of donor PFM support has a more significant and stronger association (the coefficient doubles in size) with scores for *de jure* and concentrated PFM processes, again highlighting how donor PFM support seems to focus more on rules, procedures and specific actors within government. The results are somewhat reversed when it comes to upstream vs. downstream processes. Here, the association is stronger with downstream processes. The best explanation we can offer is that this could be related to the large amounts of funding devoted to IFMIS implementation, a typical downstream PFM reform.

**Table 6. Partial regression results
(PEFA scores by type of PFM process)**

Dependent Variable PEFA Avg. Scores by PFM process	Up-stream	Down-stream	De jure	De facto	Conc. actors	Deconc. actors
PFM support	0.01	0.02***	0.02**	0.01**	0.02***	0.01*
GBS as % of total ODA	1.42	0.38	0.77	0.50	0.46	0.73
Length of donor engagement	0.02**	0.007	0.02***	0.002	0.02***	0.002
Observations	93	93	93	93	93	93
R-squared	0.52	0.70	0.64	0.69	0.65	0.67

Notes: Ordinary Least Squares regressions with robust standard errors. Only coefficients are reported.

All regressions include all other variables from main model in Table 4.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

In summary, the PEFA large-N analysis considered a number of economic/social, political/ institutional, and aid-related variables that could explain differences in the quality of PFM systems across countries. It found that income levels, economic performance (measured through average growth rates), population size and donor PFM support are significantly and positively correlated with better PFM systems, while state fragility negatively affects PFM quality. These results are robust to a series of tests and changes in econometric models. Moreover, when the analysis focuses on more recent years or on low-income countries only, the share of GBS in total aid also becomes significant in explaining differences in the quality of PFM systems, with a positive coefficient. Finally, donor support, both in terms of financial resources and of length of engagement, seems to affect different types of PFM processes differently. One of the more interesting findings for the purposes of the evaluation that this study is part of, that of the positive correlation between donor PFM aid and quality of PFM systems in recipient countries, suffers from serious endogeneity issues that we were not able to resolve, and that will need to be addressed by utilising other approaches and techniques, including qualitative case studies that can clarify and unpack the existing causal linkages through a ‘process tracing’ methodology (George and Bennett 2005).

Results of the analysis: HIPC medium-N sample

A proper assessment of the effectiveness and impact of donor support to PFM reforms would require data that capture changes in the quality of PFM systems over time. One of the main limitations of the analysis carried out so far based on PEFA data only is that it compares across countries at a certain point in time, and therefore on the assumption that better quality PFM systems today reflect successful past PFM reforms. Moreover, it suffers from a series of statistical problems that limit the scope of the analysis and the validity of the results. We therefore complement the PEFA large-N analysis with a non-econometric analysis of 19 African HIPC countries for which time-series data is available, though with more limited coverage of PFM systems (see Appendix 4), and somewhat less reliable indicators. Nevertheless, table 1 above presented summary scores for the period 2001–2010, based on information drawn from both HIPC and PEFA assessments.

Table 7 below links changes in the quality of PFM systems to a set of possible explanatory variables. It ranks countries according to whether they saw a clear improvement in the quality of their PFM systems over the period considered (+), manifested an unclear trend (=) or experienced a worsening level of PFM quality. It also shows the overall difference in summary scores between 2001 and 2010. Data for various explanatory variables are presented alongside HIPC/PEFA scores to partly replicate the model utilised for the PEFA large-N analysis, this time covering a longer period of time, from 1995 to 2008²⁹.

29 Resource dependency was not included as only one country (Cameroon) is considered resource-dependent. Regional dummies were not relevant as all countries included are in Sub-Saharan Africa. Donor PFM support is presented as the average of yearly per capita PFM support over the period 1995–2008, to take country size into account.

Table 7. The determinants of the quality of PFM systems in 19 African HIPCs, 2001–2010

Country	2001 – 2010 Variation		Avg GDP pc	Avg Growth %	Admin. Herit. ^a	Avg De-moc-racy FH index	Avg Aid Dep. %	Total PFM aid US\$ p/c	Avg GBS/Total aid %
			US\$ PPP	%				US\$ p/c	%
Zambia	+	+8	1005	3.8	GBR	4.0	21.4	0.71	9.4
Burkina Faso	+	+4	878	6.0	FRA	4.2	14.4	0.06	15.0
Ethiopia	+	+4	548	6.8	N/A	4.8	12.4	0.04	4.3
Cameroon	+	+1	1741	4.0	FRA	6.1	5.7	0.04	3.1
Mali	+	+1	856	5.5	FRA	2.4	15.4	0.06	10.0
Average			1006	5.2		4.3	13.9	0.18	8.4
Ghana	=	+8	1039	5.0	GBR	2.4	10.6	0.18	11.5
Mozambique	=	+2	548	7.6	POR	3.4	28.7	0.47	15.0
Niger	=	+1	550	4.1	FRA	4.2	14.3	0.05	9.9
Guinea	=	0	940	3.7	FRA	5.5	7.7	0.12	0.3
Tanzania	=	0	879	5.6	GBR	4.0	13.8	0.32	17.5
Madagascar	=	-1	826	3.7	FRA	3.1	13.5	0.09	6.8
Malawi	=	-1	636	4.9	GBR	3.3	22.1	0.63	7.8
Senegal	=	-1	1399	4.3	FRA	3.3	8.2	0.07	2.6
Chad	=	-2	1049	7.2	FRA	5.6	11.0	0.18	5.0
São Tomé	=	-2	1383	6.8	POR	1.7	23.4	1.35	0.9
Average			925	5.3		3.6	15.3	0.35	7.7
Rwanda	-	-2	686	10.4	BEL	6.0	23.6	0.42	16.4
Uganda	-	-2	784	7.4	GBR	4.6	13.4	0.25	9.3
Gambia	-	-4	1033	4.6	GBR	5.2	13.0	0.43	1.7
Benin	-	-5	1188	4.6	FRA	2.1	10.0	0.17	8.7
Average			923	6.7		4.5	15.0	0.32	9.0

Note: All variables refer to 1995–2008, except for GBS/Total aid (2002–06).

a) GBR=Great Britain; FRA=France; POR=Portugal; BEL=Belgium.

What the results show is that variables that are significantly correlated with higher PFM quality (measured through PEFA scores) do not

seem to explain much of the changes over time in PFM quality measured through the HIPC/PEFA indicators. Countries that saw a clear improvement do have a higher average income level and lower aid dependence than countries that either worsened or showed an unclear trend, but they are predominantly former French colonies, have a lower average growth rate and did not receive higher amounts of either PFM support or of general budget support. If we consider only the difference in score between 2001 and 2010, rather than the scores in the various assessments, the picture again changes. In this case, countries whose 2010 score is higher than the 2001 did receive higher amounts of general budget support, but have lower average incomes, economic growth rates and per capita PFM support.

Moreover, many of these differences across country groups are not likely to be significant, given the high levels of within-group variance. The better-performing group includes countries with income levels that range from US\$ 1741 (Cameroon) to US\$ 548 (Ethiopia), and with levels of per capita donor PFM support that varies between US\$ 0.71 (Zambia) and US\$ 0.04 (Ethiopia and Cameroon) a year. Similarly, the group of countries who saw a decline in the quality of their PFM systems includes both Rwanda and Gambia, who received one of the highest and one of the lowest levels of general budget support as a share of total aid, and Uganda and Benin, which score very differently with regard to the strength of their democratic institutions.

In other words, what the HIPC medium-N analysis shows is that the results from the PEFA large-N analysis need to be taken with some caution. Some of the differences that exist between present PFM quality and reform history need to be taken into account when interpreting results, even though the much smaller sample clearly does not permit to carry out a similarly detailed analysis. Similarly, the lack of more detailed information on PFM support in donor responses to our data request did not permit a more in-depth analysis of how specific PFM-related activities funded by donors may have influenced PFM reform outcomes, either in general or in specific PFM areas. The best that the HIPC medium-N analysis can do given the existing constraints, as we argue in the conclusions, is to guide the choice of case study countries that can provide more detailed comparative insights into the dynamics of PFM reforms, in order to better explain their successes and failures.

Conclusions and implications for overall evaluation

This study brought together available quantitative evidence on the quality of PFM systems across a large number of countries, with the aim of assessing the factors that are associated with and may have determined cross-country differences and variations over time. In order to operationalise and measure the quality of PFM systems, we drew from two main sources: a cross-country dataset of PEFA assessments carried out worldwide in more than 100 countries, and a more limited panel dataset based on a combination of HIPC and PEFA indicators, covering 19 African countries over the period 2001 to 2010. A specific aim of the study was to investigate the effectiveness and impact of donor PFM support on the quality of PFM systems in recipient countries. For this purpose, and given the scarcity and lack of reliability of existing data, we devoted substantial resources to the compilation of a new dataset gathering information on donor activities in support of PFM reforms over the period 1995 to 2008. Finally, we included in the analysis a series of economic/social, political/institutional and aid-related variables that were identified as having a potential influence on PFM quality and reforms.

The first phase of the analysis focused on the use of econometric techniques to identify the key variables which were significantly associated with changes in the quality of PFM systems. Results largely confirmed some of the findings of previous research. Among the different sets of factors considered, economic and social ones explain more than half of the existing differences in the quality of PFM systems, while aid-related factors explain only a quarter of the variation. Countries with higher income levels, good recent economic performance, and larger populations have better average PEFA scores. State fragility, defined as countries being in a conflict or post-conflict situation and measured through the presence of peacekeeping forces, is negatively associated with the quality of PFM systems.

More interestingly, the analysis found a significant and positive association, albeit with a very small coefficient, between donor PFM

support and average PEFA scores. If more recent data are used, or when the analysis focuses on low-income countries only, the share of total aid provided as general budget support is also positively associated with better PFM quality. These results survived a number of robustness checks. Moreover, if looking at both financial resources and length of engagement with PFM reforms, higher donor support is more strongly associated with improvements in *de jure* and concentrated PFM processes. Substantive investments in IFMIS systems, however, seem to have had a positive (albeit limited) impact on the quality of PEFA dimensions related to budget execution.

The impact of donor PFM support on PEFA scores cannot be proven, however, given that the cross-sectional nature of available data does not exclude reverse causation, or the possibility that it could be countries with higher PEFA scores that receive more PFM-related assistance, rather than the other way around. Results may also reflect biases in the data themselves, given limitations in the data collection process. In this sense, improvements in PEFA scores may in fact bear little relation with donor PFM support, and may reflect instead isomorphic tendencies, a hypothesis supported by the fact that *de jure* PEFA dimensions score consistently better than *de facto* ones.

Various attempts were made at finding adequate instrumental variables that could untangle the direction of causation, but without success. The problem of endogeneity has been widely noted and discussed in the literature linking foreign aid and institutions³⁰. Acemoglu (2005) is generally pessimistic about the possibility of identifying causal relationships in comparative political economy, and argues that robust non-causal relationships nevertheless are of value to advancements in theoretical analysis and policymaking. In our case, as argued below, the positive and significant association between PEFA scores and donor PFM support provides an interesting basis for the country case studies that will complement the analysis carried out in this paper.

Results from the second phase of the analysis, looking at changes in HIPC/PEFA indicators over time, only partly support the find-

30 See Knack (2001), Svensson (2000), Tavares (2003), Alesina and Weder (2002), Brautigam and Knack (2004), Rajan and Subramanian (2007), and Ear (2007).

ings mentioned above, and highlight the fact that the characteristics of African HIPC countries that were successful in implementing PFM reforms vary substantially, even across the few variables that were found to be significant in the first phase of the analysis. A more detailed look at the influence of specific types of donor PFM support on PFM systems was unfortunately prevented by a lack of sufficient data provided by donor agencies.

In fact, probably the best way to gain a more detailed understanding of the factors affecting PFM reforms, and to address the difficulty of confirming the direction of causality between donor PFM support and the quality of PFM systems, is to complement the present study with in-depth qualitative case studies that can unearth the underlying dynamics of PFM reforms and the role that donors played in supporting them.

One of the key issues that this study can help address, therefore, is the choice of countries that should be the focus of the case study approach. Both the PEFA large-N and the HIPC medium-N analyses can assist in this regard. As suggested by Lieberman (2005), the results of a large-N analysis can inform the choice of in-depth case studies through the analysis of regression residuals, or the differences between the real values of the dependent variable (average PEFA scores) and those generated by the regression. Cases should be chosen either within those with very small residuals, in the attempt to confirm the findings from the regression analysis, or within the group of countries with larger residuals, to open the investigation to possible alternative models and explanations (see Figure 5 for a plot of residuals). Countries that belong to the first group include Afghanistan, Cambodia, Ghana, Honduras, Papua New Guinea and Sierra Leone. Belize, India, Malawi, Tunisia and Yemen belong instead to the latter group.

and on governments' own efforts to strengthen PFM systems), to further test some of the findings presented above. Qualitative case studies will be fundamental, too, to complement and address the many shortcomings of quantitative analysis, most notably the difficulties in explaining not only *if* and *when* donor PFM support has had an impact on PFM systems, but also *why* and *how* it did, taking into account the differences in country context in which PFM reforms take place.

References

- Acemoglu, D. (2005) 'Constitutions, Politics, and Economics: A Review Essay on Persson and Tabellini's *The Economic Effects of Constitutions*'. *Journal of Economic Literature*, 43(4), 1025–1048.
- Alesina, A. and B. Weder (2002) 'Do Corrupt Governments Receive Less Foreign Aid?' *American Economic Review*, 92(4), 1126–1137.
- Andrews, M. (2010) *How Far Have Public Financial Management Reforms Come in Africa?* RWP10–018 Faculty Research Working Paper Series. Cambridge, MA: Harvard Kennedy School.
- Andrews, M. (2009) 'Good Government Means Different Things in Different Countries', *Governance*, 23(1): 7–35.
- Andrews, M. (2007) "What would an ideal PFM system look like?" in A. Shah (ed.) *Budgeting and Budgetary Institutions*. Washington, DC: World Bank.
- Arndt, C. (2008) 'The Politics of Governance Ratings', *International Public Management Journal*, 11(3), 275–297.
- Brautigam, D. and S. Knack (2004) 'Foreign aid, institutions, and governance in sub-Saharan Africa'. *Economic development and cultural change*, 52(2), 255–286.
- Busse, M. et al. (2007) *Institutions, Governance and Trade*. Hamburg: Hamburg Institute of International Economics (HWWI).
- Cheibub, J.A., J. Gandhi and J.R. Vreeland (2010) 'Democracy and Dictatorship Revisited', *Public Choice*, 143: 67–101.
- de Renzio, P. (2009a) *Taking Stock: What do PEFA Assessments tell us about PFM systems across countries?* ODI Working Paper 302. London: Overseas Development Institute.
- de Renzio, P. (2009b) *Assessing Budget Institutions and Budget Reforms in Developing Countries: Overview of theoretical approaches and empirical evidence*. Approach & Methodology for the Evaluation of Donor Support to Public Finance Management (PFM) Reform in Developing Countries. Part A. Oxford: Fiscus
- de Renzio, P., P. Gomez and J. Sheppard (2009) 'Budget Transparency and Development in Resource-Dependent Countries', *International Social Science Journal*, 57: 57–69.

- de Renzio, P. and W. Dorotinsky (2007) *Tracking Progress in the Quality of PFM Systems in HIPCs: An update on past assessments using PEFA data*. Washington, DC, PEFA Secretariat.
- Ear, S. (2007) 'Does Aid Dependence Worsen Governance?' *International Public Management Journal*, 10(3), 259–286.
- Fagernas, S. and J. Roberts (2004) *Fiscal Impact of Aid: A Survey of Issues and Synthesis of Country Studies of Malawi, Uganda and Zambia*. ESAU Working Paper 11. London: Overseas Development Institute.
- George, A. L. and A. Bennett (2005). *Case studies and theory development in the social sciences*. Cambridge, MA: MIT Press.
- Hedger, E. and P. de Renzio (2010) *What do Public Financial Management assessments tell us about PFM reform?* ODI Background Note. London: Overseas Development Institute.
- IDA/IMF (2005) *Update on the Assessments and Implementation of Action Plans to Strengthen Capacity of HIPCs to Track Poverty-Reducing Public Spending*. Washington DC: International Development Association and International Monetary Fund.
- IDA/IMF (2003) *Country Assessment and Action Plan (AAP) for HIPCs: Questionnaire, Benchmarks, Explanations, Standard Tables*. Washington DC: International Development Association and International Monetary Fund.
- IMF (2010) *Budget Institutions and Fiscal Performance in Low-Income Countries*. IMF Working Paper WP/10/80. Washington DC: International Monetary Fund.
- IMF (2007b). *Fiscal Policy Response to Scaled-Up Aid: Strengthening Public Financial Management*. Washington, DC: International Monetary Fund.
- Kakabadse, A., N.K. Kakabadse and A. Kouzmin (2003) 'Reinventing the Democratic Governance Project through Information Technology? A Growing Agenda for Debate', *Public Administration Review*, 63(1): 44–60.
- Kaufmann, D., A. Kraay and P. Zoido-Lobaton (1999) *Governance Matters*. Policy Research Working Paper 2196. Washington DC: World Bank.
- Knack, S. (2001) 'Aid Dependence and the Quality of Governance: Cross-Country Empirical Tests'. *Southern Economic Journal*, 68(2), 310–329.

- Knack, S. and A. Rahman (2007) 'Donor Fragmentation and Bureaucratic Quality'. *Journal of Development Economics*, 83, 176–197.
- La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny (1999) 'The Quality of Government' *Journal of Law, Economics, and Organization*, 15(1): 222–279.
- Lieberman, E. S. (2005) "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review*, 99(3): 435–452.
- Moss, T., G. Pettersson, et al. (2006). *An Aid-Institutions Paradox? A Review Essay on Aid Dependency and State Building in Sub-Saharan Africa*. Working Paper No. 74. Washington, DC: Centre for Global Development.
- Munck, G. (2009) *Measuring Democracy: A Bridge Between Scholarship and Politics*. Baltimore, MD: Johns Hopkins University Press.
- Pack, H. and J.R. Pack (1990) 'Is Foreign Aid Fungible? The Case of Indonesia' *The Economic Journal*, 100(399), 188–194.
- PEFA (2009) *Issues in Comparison and Aggregation of PEFA Assessment Results Over Time and Across Countries*. Washington, DC: Public Expenditure and Financial Accountability (PEFA) Secretariat.
- PEFA (2005). *Public Financial Management Performance Measurement Framework*. Washington, DC: Public Expenditure and Financial Accountability (PEFA) Secretariat.
- Rabe-Hesketh, S. and B. S. Everitt (2007) *A Handbook of Statistical Analyses Using Stata*. 4th edition. Boca Raton, FL: Chapman and Hall.
- Rajan, R. and A. Subramanian (2007) 'Does Aid Affect Governance?' *American Economic Review*, 97(2), 322–327;
- Remmer, K. L. (2004) 'Does Foreign Aid Promote the Expansion of Government?' *American Journal of Political Science*, 48(1), 77–92.
- Rivera-Batiz, F. L. (2002) 'Democracy, Governance and Economic Growth: Theory and Evidence', *Review of Development Economics*, 6(2): 225–247.
- Svensson, J. (2000) 'Foreign Aid and Rent-Seeking', *Journal of International Economics*, 51, 437–461.
- Tavares, J. (2003) 'Does Foreign Aid Corrupt?' *Economics Letters*, 79, 99–106.
- Treisman, D. (2000) 'The Causes of Corruption: A Cross-National Study', *Journal of Public Economics*, 76: 399–457.

- Vyas, S. and L. Kumaranayake (2006) “Constructing socio-economic status indices: how to use principal components analysis.” *Health Policy and Planning*, 21(6): 459–468.
- Wehner, J. and P. de Renzio (2010) *Citizens, legislators, and executive disclosure: The political determinants of fiscal transparency*. Unpublished manuscript. London: London School of Economics.
- Wescott, C. G. (2009). “World Bank Support for Public Financial Management and Procurement: From Theory to Practice.” *Governance*, 22(1): 139–153.
- World Bank (2008). *Public Sector Reform: What Works and Why? An IEG Evaluation of World Bank Support*. Washington, DC: World Bank, Independent Evaluation Group.

Appendix 1: List of PEFA Countries

	Country	Date	Status		Country	Date	Status
1	Afghanistan	Jun. 08	Final-P	51	Lesotho	Jul. 09	Draft
2	Albania	Jul. 06	Final-P	52	Liberia	Jun. 09	Final
3	Armenia	Oct. 08	Final-P	53	Madagascar	May 08	Final
4	Azerbaijan	Jan. 08	Final	54	Malawi	Jun. 08	Final
5	Barbados	Oct. 06	Draft	55	Maldives	Nov. 09	Final
6	Belarus	Apr. 09	Final	56	Mali	Dec. 08	Final-P
7	Belize	Oct. 08	Draft	57	Mauritania	Mar. 08	Final
8	Benin	Sep. 07	Final-P	58	Mauritius	Jun. 07	Final-P
9	Bolivia	Aug. 09	Draft	59	Moldova	Jul. 08	Final-P
10	Botswana	Feb. 09	Final-P	60	Montenegro	Jul. 09	Final
11	Brazil	Oct. 09	Draft	61	Morocco	May 09	Final-P
12	Burkina Faso	Apr. 07	Final-P	62	Mozambique	Feb. 08	Final-P
13	Burundi	Feb. 09	Final	63	Namibia	Nov. 08	Final
14	Cambodia	Mar. 10	Draft	64	Nepal	Feb. 08	Final-P
15	Cameroon	Jan. 08	Final	65	Niger	Dec. 08	Draft
16	Cape Verde	Dec. 08	Final-P	66	Pakistan	Jun. 09	Final
17	Central African Republic	Jun. 08	Final	67	Papua New Guinea	Mar. 09	Draft
18	Chad	Jul. 09	Final	68	Paraguay	Apr. 08	Final-P
19	Colombia	Jun. 09	Draft	69	Peru	Apr. 09	Final-P
20	Comoros	Jan. 08	Final	70	Philippines	Oct. 07	Draft
21	Congo, Dem. Republic of	Mar. 08	Final	71	Russian Federation	Jan. 07	Draft
22	Congo, Republic of	Mar. 06	Final-P	72	Rwanda	Jun. 08	Final-P
23	Cote d'Ivoire	Nov. 08	Final-P	73	Samoa	Oct. 06	Final-P
24	Dominica	Apr 07	Draft	74	S. Tome and Principe	Jan. 10	Final
25	Dominican Republic	Nov. 09	Final	75	Senegal	Dec. 07	Final

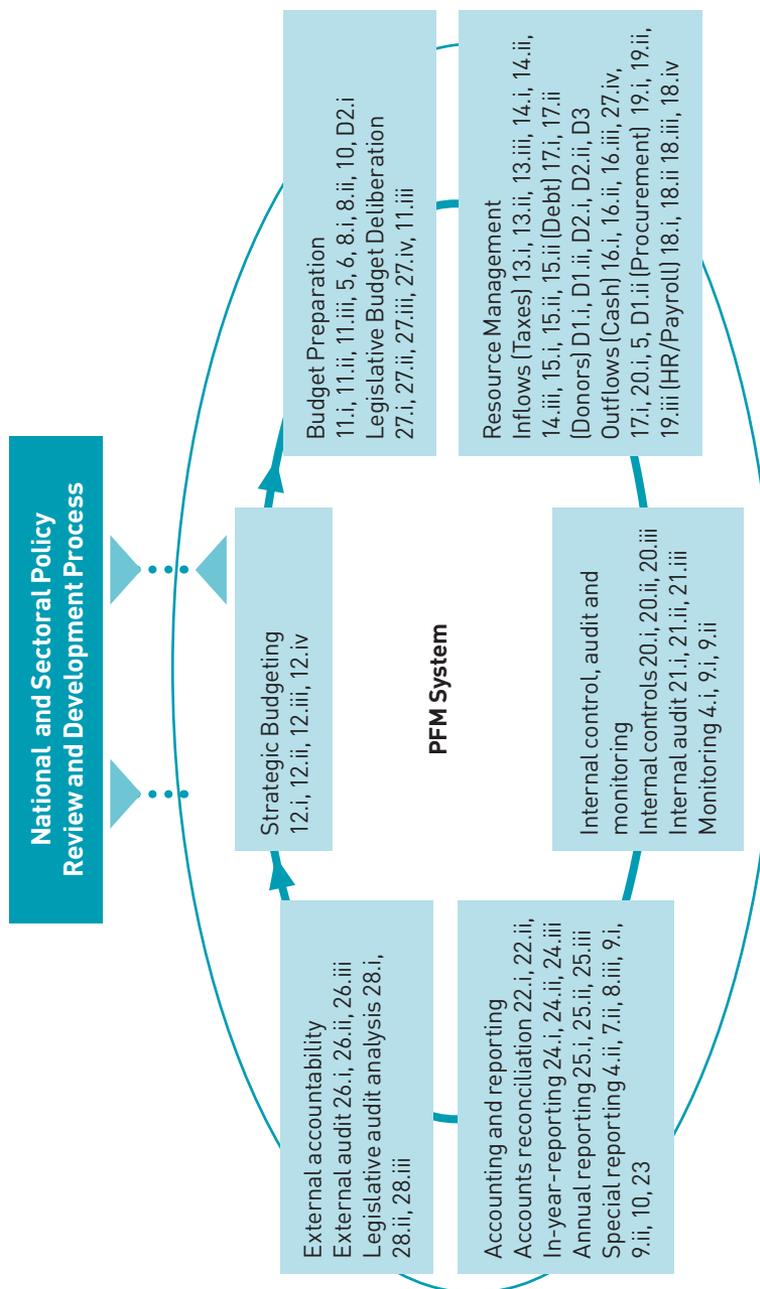
Country	Date	Status	Country	Date	Status		
26	Egypt	May, 09	Draft	76	Serbia	Feb. 07	Final-P
27	El Salvador	May 09	Final-P	77	Seychelles	Dec. 08	Final
28	Ethiopia	Oct. 07	Final-P	78	Sierra Leone	Dec. 07	Final-P
29	Fiji Islands	Jun. 05	Final	79	Solomon Islands	Nov. 08	Final-P
30	FYR Macedonia	Aug. 07	Final-P	80	South Africa	Sep. 08	Final-P
31	Gambia	Mar. 09	Draft	81	St. Kitts and Nevis	Sep. 09	Draft
32	Georgia	Nov. 08	Final-P	82	St. Lucia	Nov. 09	Draft
33	Ghana	Jan. 10	Final	83	St. Vincent and Gren.	Sep. 06	Final
34	Grenada	Mar. 10	Final-P	84	Sudan	May 09	Draft
35	Guatemala	Mar. 10	Draft	85	Swaziland	Mar. 10	Draft
36	Guinea	Jul. 07	Final	86	Syria	Mar. 06	Final
37	Guinea Bissau	May 09	Final	87	Tajikistan	Jun. 07	Final-P
38	Guyana	Dec. 07	Draft	88	Tanzania	Jun. 09	Final
39	Haiti	Jan. 08	Final-P	89	Thailand	Oct. 09	Final
40	Honduras	Apr. 09	Final	90	Timor Leste	Feb. 07	Final-P
41	India	Mar. 10	Final-P	91	Togo	Nov. 08	Draft
42	Indonesia	Oct. 07	Final-P	92	Tonga	Sep. 07	Final
43	Iraq	Jun. 08	Final	93	Trinidad and Tobago	Dec. 08	Final-P
44	Jamaica	May 07	Final-P	94	Tunisia	Mar. 10	Draft
45	Jordan	Apr. 07	Final-P	95	Turkey	Dec. 09	Final
46	Kazakhstan	Jun. 09	Final	96	Uganda	Jun. 09	Final-P
47	Kenya	Mar. 09	Final-P	97	Ukraine	Mar. 07	Final-P
48	Kiribati	Dec. 09	Draft	98	Vanuatu	Nov. 09	Final
49	Kyrgyz Republic	Oct. 09	Final	99	Yemen	Jun. 08	Final-P
50	Lao PDR	Dec. 09	Draft	100	Zambia	Jun. 08	Final

Note: 'Final-P' denotes reports that have been made public on the PEFA website.

Appendix 2: List of HIPC/PEFA Countries

	Country	HIPC Assessments	PEFA Assessments
1	Benin	2001, 2004	2007
2	Burkina Faso	2001, 2004	2007
3	Cameroon	2001, 2004	2008
4	Chad	2001, 2004	2009
5	Ethiopia	2001, 2004	2007
6	Gambia	2001, 2004	2009
7	Ghana	2001, 2004	2006, 2010
8	Guinea	2001, 2004	2007
9	Madagascar	2001, 2004	2006, 2008
10	Malawi	2001, 2004	2006, 2008
11	Mali	2001, 2004	2008
12	Mozambique	2001, 2004	2006, 2008
13	Niger	2001, 2004	2008
14	Rwanda	2001, 2004	2008
15	São Tome & Principe	2001, 2004	2007, 2010
16	Senegal	2001, 2004	2007
17	Tanzania	2001, 2004	2006, 2009
18	Uganda	2001, 2004	2006, 2009
19	Zambia	2001, 2004	2005, 2008

Appendix 3: PFM Clusters



Detailed description of the six PFM clusters used in the analysis, with PEFA indicator dimensions included in each.

Source: Andrews (2010)

Appendix 4: HIPC/PEFA indicators

The table below explains how PEFA indicators and information was used to update HIPC indicator scores for more recent years.

	HIPC Indicator	PEFA ^a	Notes
Formulation	1. Coverage of the budget or fiscal reporting entity	Intro PI-8 PI-9	The HIPC indicator is not directly related to any one PEFA indicator, but information is often available in different parts of the PEFA report.
	2. Degree of spending being funded by inadequately reported extra-budgetary sources	PI-7 (i)	Directly related. Conversion sometimes difficult due to differences in thresholds.
	3. Reliability of the budget as a guide to future outturn	PI-1 PI-2	Directly related and easily convertible.
	4. Inclusion of donor funds	PI-7 (ii)	Directly related and easily convertible.
	5. Classification	PI-5	Directly related and easily convertible.
	6. Identification of poverty-reducing spending		Information not available in PEFA reports. Indicator dropped.
	7. Integration of medium-term forecasts	PI-12	Directly related and easily convertible.
Execution	8. Evidence of budget execution problems – arrears	PI-4	Directly related. Limited information available in some PEFA reports.
	9. Effectiveness of internal control system	PI-18 PI-20 PI-21	Easily convertible based on the information available for different indicators.
	10. Tracking surveys are in use	PI-23	Directly related and easily convertible.
	11. Quality of fiscal information	PI-22 (i)	Directly related and easily convertible.

Reporting	12. Regularity of timely internal fiscal reporting		Information not available in PEFA reports. Indicator dropped.
	13. Regular fiscal reports track poverty spending		Information not available in PEFA reports. Indicator dropped.
	14. Transactions are recorded in the accounts in a timely fashion		Information not available in PEFA reports. Indicator dropped.
	15. Timeliness of audited financial information	PI-26	Directly related and easily convertible.
Procurement	16. Efficiency and effectiveness of the public procurement system	PI-19	Indicator dropped because (a) not included in 2001 HIPC assessment, and (b) PEFA indicator looks at different issues.

- a) Indicates the section of the PEFA assessment report or the indicator (and indicator dimension) in the PEFA Performance Measurement Framework. Indicators 1, 2, 4 and 5 were utilised to generate the *transparency and comprehensiveness* (INFO) score; indicators 3, 7 and 10 were utilised to generate the *linking budgeting, planning and policy* (POL) score; indicators 8, 9, 11 and 15 were utilised to generate the *control, oversight and accountability* (CTRL) score. Letter scores were transformed into numerical ones (C=1, B=2, A=3) and added together in each cluster.

Appendix 5: Donor data request form

1. Cover page



Request for Data on Donor Support to Public Financial Management (PFM) Reforms

As part of an evaluation of donor support to public financial management (PFM) reform in developing countries the Overseas Development Institute (ODI) is carrying out an analytical study of quantitative cross-country evidence. The joint evaluation was promoted by the evaluations departments of DANIDA, Sida, DFID and the African Development Bank, in consultation with the OECD-DAC Evaluation Network.

A key part of this will be the analysis of donor support to PFM reforms. To this end we kindly request that your agency provides us with information on your support to PFM reform. This includes Technical Assistance projects, training/capacity building, and PFM-related analytical work (e.g. PEFAs, PERs) with a minimum value of US\$ 100,000, disbursed over the period 1995-2008. It does not include, for example, administrative costs of any sorts, the cost of seconded staff, and budget support operations (even though these are classified as PFM support by some agencies). If no yearly data is available, disbursements for multi-year projects should be recorded in the final year of implementation. For ongoing projects, best available estimates for disbursements up to 2008 are sufficient.

A first set of data is requested for the 100+ countries where a PEFA Assessment has been carried out (sheet 1). For these countries, we simply need data on funds disbursed for PFM-related activities over the period 1995-2008.

A second, more detailed set of data is requested for the 19 African countries where both HIPC and PEFA Assessments were carried out (sheet 2). For these countries, we need more detailed information about each project/activity, including some details about the kind of assistance provided and the PFM areas covered. Copies of specific project documents, where available, would also be greatly appreciated.

We are available to support you in the collection of this data. If needed, one of our researchers can also visit your office and assist directly. Please contact us for any query via Paolo de Renzio (p.derenzio.ra@odi.org.uk) or Ed Hedger (e.hedger@odi.org.uk).

Thank you for your collaboration!

2. PEFA large-N sample

Support to PFM reforms (disbursements in current US\$ or local currency, please state)															
Country	Region	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Country 1															
Country 2															
Country 3															
...															

3. HIPC medium-N sample

NB each project should be on one line, if there are multiple projects for a country then please add in lines as required.

Project Details		Disbursements (current US\$ or local currency, please state)														
Project title	Start year	End year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Country 1																
Country 2																
Country 3																
...																

a) type of assistance provided (e.g. TA, training, equipment, research, etc.)
 b) main phase of the budget cycle concerned (formulation, approval, execution, audit)
 c) Area or focus of assistance (e.g. MTEF performance-based budgeting, transparency, audit IFMIS, revenue administration, etc.)

Appendix 6: List of variables and summary statistics

Variable	Description	Source
Economic/social variables		
GDP per capita	Per capita Gross Domestic Product (US\$ PPP), average 2002–06	World Development Indicators
Recent GDP growth	Year-on-year Growth in Gross Domestic Product (%), average 2002–06	World Development Indicators
Aid dependency	Official Development Assistance as a share of Gross National Income (%), average 2002–06	World Development Indicators
Resource dependency	Dummy variable for countries which rely heavily on oil and mineral revenues	International Monetary Fund
Population	Total Population, average 2002–06	World Development Indicators
Adult literacy	Adult literacy rates (%), average 2002–06	Human Development Report
Trade openness	Sum of total imports and total exports as a share of Gross Domestic Product (%), average 2002–06	World Economic Outlook
Technological diffusion	Electric power consumption (kWh per capita) Telephone subscribers (per 100 people) Personal computers (per 100 people) Internet users (per 100 people)	World Development Indicators
Political/institutional variables		
Administrative heritage	Dummy variables for former British and French colonies	Quality of Government
Democracy	Freedom House sum of Civil Liberties and Political Rights Index (1=most democratic, 7=least democratic)	Freedom House
State fragility	Dummy variable for presence of peacekeepers in the country since 1995	[Data obtained from Matt Andrews]
Regional dummies	Dummy variables for region of the world to which country belongs	World Development Indicators

Variable	Description	Source
Aid-related variables		
PFM support	Donor support for PFM reforms	Own data collection
Aid fragmentation	Fragmentation index calculated on the basis of donor shares of total ODA received by country, average 2002–06	OECD/DAC database
GBS as % of total ODA	General Budget Support as a share of total ODA received by country (%), average 2002–06	OECD DAC/CRS database
IMF programme	Dummy variable for presence of IMF programme for at least one year during 2002–06	International Monetary Fund
WB PRSC	Dummy variable for presence of World Bank Poverty Reduction Support Credit for at least one year during 2002–06	World Bank
Length of donor engagement	No. of years since first World Bank PFM project	[Data obtained from Matt Andrews]

Variable	Obs.	Mean	Std. Dev.	Min	Max
PEFA average score	100	2.44	0.48	1.38	3.58
GDP per capita, PPP US\$	99	3,902	2,159	251	19,189
Recent GDP growth, %	100	4.96	3.23	4.60	9.33
Population (millions)	100	30.8	112.0	0.047	1,080.0
Resource dependency	100	0.17	0.38	0	1
Adult literacy, %	99	75.22	21.37	26.14	99.00
Trade openness	94	84.87	34.40	27.04	190.36
Technological diffusion					
Electricity (kWh pc)	100	730.93	1199.99	0	5666.81
Phone (%)	100	29.73	29.43	0	122.84
Computer (%)	100	3.44	4.14	0	18.00
Internet (%)	100	6.57	7.85	0	37.83
Democracy (Freedom House)	100	3.72	1.57	1	7
State fragility	100	0.16	0.37	0	1
Aid dependency, %	98	9.19	10.72	0	43.10
PFM support, US\$m/year	100	3.07	5.59	0	33.21
Aid fragmentation	99	0.80	0.10	0.36	0.91

Variable	Obs.	Mean	Std. Dev.	Min	Max
GBS as % of total ODA	99	0.04	0.05	0	0.30
IMF programme	100	0.58	0.50	0	1
WB PRSC	100	0.20	0.40	0	1
Length of donor engagement (years)	100	12.04	6.69	-1	22

Appendix 7: TERMS OF REFERENCE

Evaluation of Donor Support to Public Finance Management (PFM)

Reform in Developing Countries

Analytical study of quantitative cross-country evidence

CONTEXT

Larger flows of development assistance in the form of budget support are being used to increase both the volume and effectiveness of public spending in developing countries in ways that use country systems for public financial management. The increasing use of budget support as an aid modality has broadened the range of development agencies providing financial and technical assistance for public financial management (PFM).reform and, as a consequence, there is now broad international experience with different types of approach to PFM reform.

Given this emerging experience, the evaluation departments of DANIDA, Sida, DFID and the African Development Bank have agreed, in consultation with the OECD-DAC Evaluation Network, to manage a joint evaluation of public financial management reforms in developing countries. The evaluation aims to answer two questions:

- (a) Where and why do PFM reform efforts succeed?
- (b) Where and how does external support to PFM reform-efforts contribute most effectively to their success?

The evaluation will start with an analytical study that will collate cross-country data on PFM systems world wide and time series data on a smaller group of countries where comparable data exists. The study will need to investigate what data sources are available on PFM and whether these can be used to perform statistical tests and simple correlations to explain patterns and emerging trends in successful PFM reform.

The initiative to evaluate PFM reforms includes this analytical study as well as a number of other sub-projects:

1. Review of Public Financial Management Reform Literature (Pretorius and Pretorius 2009, Evaluation Report EV698, DfID)
2. Approach & Methodology for the Evaluation of Donor Support to Public Finance Management (PFM) Reform in Developing Countries (Lawson and de Renzio, 2009, Draft)
3. Analytical study of quantitative cross-country evidence (i.e. the study that this Terms of Reference refers to)
4. Africa Region Study (a consultancy study including five country case studies and a synthesis of findings – expected to be performed during second half of 2010)
5. Europe Region Study (still at early planning stage, Sida lead)
6. Possibly additional regional studies for Asia and Latin America

The quantitative study will inform the subsequent regional studies but will also, in itself, be a valuable piece of information on PFM reform progress and support to PFM reforms. The objective of the study is to bring together international and cross country data on PFM systems and support to PFM reform and see if it can be used to help support the following hypotheses:

- That the type of PFM reform is important for strengthening country systems, e.g., PFM reform is likely to be more sustainable if it based on country context and technical capacity; puts the basics³¹ in place first; considers the prioritization and sequencing of reforms, links to public administration reform, and builds human capacity. In contrast, PFM reform that imports PFM technologies and ideas from more advanced economies is less likely to be sustained.
- That the type of aid delivered is important for strengthening country PFM systems e.g., PFM systems improve where development is country'- led, aid is increasingly harmonised and aligned, and delivered through country public financial management systems³¹. Conversely, aid that continues to use “off budget” and

31 A rationale for using country PFM systems is set out in a “Report on the use of Country Systems in Public Financial Management” OECD/DAC 2008. Using PFM systems aligns aid with country policy priorities, policy making and policy implementation processes; aligns incentives; supports sound budgeting; enhances sustainability of results, reduces costs and facilitates harmonisation.

parallel systems for its delivery undermines domestic financial management and reform;

- That political economy can constrain or drive PFM reforms L e., that high level support for reform is key to its success; and
- That reform focused on accountability (to the public), transparency (of decision making and widely accessible financial information) and participation (of beneficiaries and users of public services) has a longer term impact (than other reforms) on PFM performance, through increasing the demand for ongoing PFM reform.

TASK DESCRIPTION

The study is concerned with the compilation of data on the quality of public financial management systems and donor support for PFM reform. The tasks will involve:

1. The compilation of PFM data world Wide in order to conduct cross-country analysis, which might shed light on the hypotheses outlined above. The analysis should highlight other factors which appear to impact on PFM performance e.g., GDP/capita, GDP growth, conflict and fragility, sources of revenue, levels of literacy and numeracy, number of chartered accountants in country and so on. Other linkages and points of commonality between countries should also be explored e.g., those that receive substantial donor support for PFM reform, compared with those that don't. The task would involve examining and commenting on a range of data that are used to assess the quality of PFM and budget institutions in developing countries e.g. the IMF's fiscal ROSCs, the CPIA, PEFA and other existing data sets.
2. A closer focus and more detailed analysis on those countries that have completed 2 PEFA assessments that have been accepted by PEFA as suitable for tracking PFM performance over time, plus HIPC data on the quality of budget systems.
3. Collecting detailed data on donor support for these countries including the amounts invested, the number of donors involved, number of years over which support has been provided, the approach taken e.g. application of an MTEF, IFMIS, sector focus etc. and the degree of TA assistance involved.

Assessing whether any relationship exists between donor support and changes in the quality of budget systems as shown by the PFM data e.g. PEFA assessments. Specifically the study would need to:

- List what type of data are available world wide, across countries and over time, and what this data reveals about the quality of PFM systems
- Consider which data sets could be used to describe the performance and quality of PFM systems world wide
- Relate the quality of PFM systems to other factors such as GDP per capita and so on
- Assess which data sets could be used to describe changes in PFM systems over time
- Relate changes in the quality of PFM systems over time to the quality and quantity of donor support and other factors.

SCOPE OF THE STUDY

The study will look at cross country and time series data that try to capture the quality of PFM processes and institutions. The first part of the study will attempt to find commonalities across groups of countries and try to explain why some countries have more developed PFM systems (or higher PEFA ratings) than others e.g. levels of literacy and so on.

However, the study is mainly concerned with looking at whether PFM reforms have led to changes (positive or negative) in the quality of PFM systems over time and where donor support has supported the process of change. This will involve looking at time series data over a sufficiently long enough period of time to adequately capture change in PFM systems. This will limit the second part of the study to those countries that have at least two PEFA assessments.

The most comprehensive attempt at constructing a framework of PFM reform and the quality of budget institutions is the PEFA framework (PEFA, 2005). However, while more than 150 PEFA assessments have been carried out over the past four years; these provide only a snap shot of the state of budget institutions across countries. Little can be said therefore on whether budget institutions have improved (with or without development assistance) unless two or more assessments have been undertaken. So far repeat assessments have been carried out in just a small number of countries, within a maximum time span of three years. Collecting data on donor sup-

port to PFM reform in countries with only one PEFA assessment would therefore make little sense as reliable and comparative data on PFM performance over the period would not be available.

Even where repeat PEFA assessments exist, the time period covered (three years) may not be sufficient to capture changes in the performance of budget institutions. To expand the period, PEFA assessments would need to be combined with comparable HIPC assessments that were carried out in 2001 and 2004. The methodology developed for the earlier HIPC assessments formed the basis for the PEFA methodology, and the degree of overlap between the two methodologies can be exploited to track 11 PFM indicators over the whole period 2001-2007.

However, the three sets of data are only available for 17 countries.³²

Given the available data (PEFA and HIPC) and the difficulties involved in compiling information on donor PFM related activities, it would therefore seem sensible to focus the quantitative study on these countries. However, there may be value in expanding the series to include all countries which have 2 completed PEFA assessments, to assess the scale of change over more recent times.

APPROACH AND CHALLENGES

The study will be based on the collection of data from secondary sources such as CPIA, PEFA and HIPC (possibly complemented by other sources such as the Open Budget Index, the Global Integrity Index, etc.) for PFM systems, and on DAC/CRS, PLAID, donor databases and reports, etc. for donor support. Data for control variables will also be collected from existing reputable sources (WDI, PolityIV, etc.).

There are a number of issues that are likely to arise and that will need to be addressed:

Access to CPIA and PEFA data is currently limited. Without additional access to these databases, the scope of the study will be greatly limited. A complementary but much more time-consuming approach would be to complement existing quantitative data with

32 These countries are Benin, Burkina Faso, Ethiopia, Ghana, Guinea, Madagascar, Malawi, Mali, Mozambique, Rwanda, S. Torne & Principe, Tanzania, Uganda, Zambia and Nicaragua, Honduras and Guyana.

the analysis of other qualitative sources that might generate quantitative measures (such as Public Expenditure, Reviews, CFAAs, CPARs or IMF Fiscal ROSCs³³), so as to generate additional data that can complement existing CPIA, PEFA and HIPC sources.

The most challenging task will be that of collecting data on donor support for PFM reforms. Relying exclusively on DAC/CRS data is not sufficient. Detailed data for the World Bank and the IMF (two of the main players in supporting PFM reforms) does not exist in an easily accessible format. Data for bilaterals is also deemed not very reliable, and in some cases not detailed enough. A decision will need to be taken in consultation with the Management Group on how to carry out this part of the data collection exercise, with the options of (a) using existing sources only; (b) focusing on large donors only; (c) asking for the joint efforts of a number of donor agencies in providing new and more detailed data.

While with sufficient data econometric analysis should be possible, it is very likely, that proving causation (rather than simply association) between donor support and successful PFM reforms will prove elusive, given potential issues of i.e. endogeneity and measurement error. Nevertheless, all possible avenues and approaches should be tried in order to get as much as possible out of the existing data. Visits to key bilateral and multilateral donors may also be required.

SPECIFIC OUTPUTS

The study will produce three main outputs:

1. A brief inception report that sets out:
 - (i) what internationally comparable (cross country and time series) data sets on PFM systems are accessible;
 - (ii) what international data are available (and useful) but not readily accessible;
 - (iii) what utility does the data have; can it be used to test basic statistical relationships and what are the proposed avenues for statistical exploration?;

33 This would require an initial exploratory phase to test whether and how much these sources can generate additional, reliable and comparable data. IMF ROSCs, given their more standardised methodology, could be more amenable to such an approach, but other diagnostic studies promoted by the World Bank also contain useful information that could be exploited for the purposes of this study.

- (iv) what data on donor support are available, in what regions should it be collected, and is it possible to explore linkages between donor support and improvements in PFM?
 - (v) tentative report outline
2. A paper of publishable quality of max 40 pages including annexes, summarising the results of the research, with a particular emphasis on (a) the findings regarding the relationship between budget institutions and the impact of donor support on PFM reforms, and (b) the quality of the data collected and utilized and related challenges.
 3. A comprehensive database in both Excel and STATA (or other statistical software) formats including all data collected and analysed related to budget institutions and outcomes, donor support for PFM reforms, and other variables of interest used in the analysis carried out in the paper.

INDICATIVE STUDY TEAM COMPOSITION

It is expected that two to three consultants will be needed, one leading the study and the other(s) providing data collection and data analysis support. The team is expected to be: knowledgeable about PFM reforms and support to PFM reforms; familiar with data relating to PFM reforms and support to PFM reforms and; be skilled in statistical/econometric methods.

TIMING

The quantitative study is expected to be finalised by end July 2010.

Study Schedule

Inception Report	2010-04-11
MG comments on draft inception re	2010-04-16
First draft report	2010-06-11
MG/RG comments on draft	2010-06-25
Final draft report	2010-07-09
MG approves/comments final draft	2010-07-16
Final report (publishable quality)	2010-07-30

ANNEX: ABBREVIATIONS

CFAA	Country Financial Accountability Assessment
CPAR	Country Procurement Assessment Report
CPIA	Country Policy & Institutional Assessment
CRS	Creditor Reporting System
HIPC	Highly Indebted Poor Countries
IFMIS	Integrated Financial Management System
MTEF	Medium Term Expenditure Framework
PEFA	Public Expenditure & Financial Accountability
PFM	Public Financial Management
ROSC (IMF)	Report on the Observance of Standards & Codes
TA	Technical Assistance
WDI	World Development Indicators

Evaluation of Donor Support to Public Financial Management (PFM) Reform in Developing Countries

Analytical study of quantitative cross-country evidence

Final Report

This study was prepared as part of a broader evaluation of donor support to public financial management (PFM) reforms. It analyses quantitative evidence on the quality of PFM systems, and assesses factors that may have determined cross-country differences and variations in the quality of PFM systems over time.

The cross country econometric analysis conducted as part of the study points to a number of findings. Countries with higher levels of per capita income, larger populations, and better economic growth are characterised by better quality PFM systems. In contrast, countries in conflict or post-conflict situations tend to have poorer quality of PFM systems. These economic and institutional factors explain most of the cross-country variation in PFM performance. The study also finds that, on average, countries that received more PFM-related technical assistance tend to have better PFM systems, but here the direction of causation could not be determined.

The study was constrained by a number of limitations and challenges, in particular poor data quality and limited information on donor PFM support. These limitations point to the need to interpret the results of the analysis with caution.



SWEDISH INTERNATIONAL DEVELOPMENT COOPERATION AGENCY

Address: SE-105 25 Stockholm, Sweden.

Visiting address: Valhallavägen 199.

Phone: +46 (0)8-698 50 00. Fax: +46 (0)8-20 88 64.

www.sida.se sida@sida.se

