POLICY RESEARCH WORKING PAPER 7084

Strengthening Public Financial Management

Exploring Drivers and Effects

Verena Fritz
Stephanie Sweet
Marijn Verhoeven

WORLD BANK GROUP

Abstract

This paper explores two relationships, first between country characteristics and the quality of public financial management ('drivers'), and second between the quality of public financial management systems and expected outcomes ('effects'). On the influence of country characteristics, the paper investigates economic factors (income level, growth, and resource dependency), population size, levels and sources of revenue, and three macro-political characteristics—political stability, regime type, and the presence of programmatic parties. These characteristics jointly explain about 40 percent of the variation in the quality of public financial management across countries. Furthermore, first-difference analysis suggests that countries with lower initial public financial management quality improve at a higher rate over time. This implies that structural factors set the scene for the likelihood of better or worse performance, but also that there is substantial variation among countries sharing certain characteristics and reform opportunities exist even in unfavorable environments. Methodologically, a key limitation is that the direction of causality cannot be fully addressed with the types of data available. On the effects of the performance of public financial management, the paper finds evidence that stronger performance results in better budget credibility, but not in lower deficits. Furthermore, there is no clear evidence regarding operational efficiency. The observed disconnect could be caused by missing complementary state capacities, measurement problems, or other issues, which need to be explored further. Overall, the findings are consistent with the assumption that stakeholder incentives and constellations matter and that reform approaches combining good technical calibration and political economy considerations are likely to influence success in strengthening public financial management.

Strengthening Public Financial Management: Exploring Drivers and Effects

Verena Fritz
Stephanie Sweet
Marijn Verhoeven[1]

---

[1] Authors' contact information: vfritz@worldbank.org, ssweet@worldbank.org and mverhoeven@worldbank.org

Contents

Acknowledgements

1. Introduction

Public financial management (PFM) systems are a core area for reform in many developing countries. PFM systems support decision making on fiscal policy and underpin budget implementation and reporting. Shortcomings in such systems can lead to lack of fiscal discipline and macro-economic instability, weaken the alignment between the allocation of public resources and national policy priorities, and contribute to greater waste and corruption in the delivery of public services.

Donors have been actively engaged in supporting the reform of PFM systems across many developing countries. However, progress with strengthening PFM systems has varied – with greater success in some areas, countries, and time periods and less so in others. This experience is leading to two complementary searches: one for technical interventions that may be a 'better fit' and better targeted at problem-solving, and a related search for better understanding of how non-technical and especially political economy factors come into play and influence where and which PFM reforms succeed.

However, if we think about tailoring PFM reforms to particular country circumstances, we have to first ascertain to what extent PFM performance is influenced by broad country characteristics. For example, if a country's income per capita and its strength of PFM systems are highly correlated and if the rate of variation is high (i.e., higher income is associated with much better PFM systems), then it follows that the scope for improving PFM systems in a low income country will tend to be limited – no matter how technically sophisticated and politically smart the reform approach is designed. Conversely, if the picture is more mixed, and statistically the relationship between income per capita and PFM performance is weak or absent, then in principle, potential progress is less constrained by income levels, and the search for specific 'good fit' and smart reform approaches is more likely to be of value. Apart from GDP per capita, there are a number of other country characteristics that have been proposed to influence the performance of PFM systems, and which we explore in this paper with a view to testing the overall limits that such characteristics might set for successful PFM reforms.

Over the past decade, comparative data about PFM systems has become available through the PEFA (Public Expenditure and Financial Accountability) initiative and the CPIA (Country Policy and Institutional Assessments). This has helped spur new avenues for research into determinants of PFM quality. Building on previous research in particular by de Renzio, Andrews and Mills 2011, we explore two relationships in this paper: (i) the relationship between country characteristics and PFM performance (both levels and changes over time), and (ii) the quality of PFM systems and expected outcomes. We do this via three main guiding questions:

1. How much of the variation in the quality of PFM systems across countries can be accounted for by macro-level country characteristics, such as income level, mineral resource dependency, and a list of other factors?

2. Can changes in the quality of PFM systems within countries over time be explained by trends in these macro-level country characteristics?

3. Does higher PFM quality translate into better fiscal outcomes and improved efficiency in public service delivery (what does available evidence show with regard to the assumed relationship)?[2]

Answers to the first two questions can inform the diagnosis of PFM shortcomings relative to other countries with similar circumstances, and it can help reform stakeholders to assess the potential scope for PFM system strengthening. Moreover, as indicated above, a weak association between characteristics and PFM status and reform progress would indicate that the former play no or a limited role in constraining reforms, and that therefore, smart reform strategies matter. The third question centers on the key rationales underlying the significant global support for PFM reforms—that this promotes better government performance and service delivery. However, efforts at empirical evidencing have been limited to date, so this remains an important issue to explore.

With regard to the first question, we find that macro-level country characteristics identified in this paper account for about 40 percent of the variation in PFM quality across countries (measured with aggregate PEFA scores). Among the characteristics explored, we find that the quality of PFM systems is most significantly and robustly associated with a country's income per capita and growth (positively) and having a high share of revenues that are obtained from natural resources (negatively). In addition, we find statistically weaker associations with the size of the overall population (positive) and being a small island developing state (negative); as well as a positive association with political stability that appears to be quite strong. With regards to the political regime and tax revenue, the statistical significance as well as the potential impact is weak; while programmatic parties appear to possibly have a substantial impact, but again, the statistical significance is borderline (at the 10 percent level). There is no association with the volume of aid relative to GNI.[3] Overall, therefore, there is some important pattern in terms of 'structural factors' setting the scene for the likelihood of better or worse PFM performance, but also considerable remaining variation among countries that share certain characteristics or even combination of characteristics.

In a second step, we employ a first differences method (i.e. analyzing *within* country changes *over* time). This hones in on the extent to which changes in country characteristics effect changes in the PFM systems, and helps diminish possible omitted variable bias (OVB) that was possible under our cross-sectional analysis. In this model, except for SIDS, we do not find evidence that country characteristics pre-determine progress on the performance of PFM. Initial level of PFM quality is the only other statistically significant variable, with a negative sign. This indicates that countries with worse initial PFM systems have tended to achieve a greater degree of reforms. It must be noted, however, that the number of observations for the first-difference

---

[2] In particular, we build on earlier work by: (i) adding more recent data and testing additional variables of country characteristics (ii) considering CPIA-13 scores as a measures of PFM quality in addition to PEFA scores, (iii) addressing the second question by supplementing the investigation of cross-country differences in the *level* of PFM system quality with a preliminary econometric assessment of whether *changes* in country characteristics are associated with *changes* in PFM performance, and (iv) adding an exploration of the effects of PFM reforms.

[3] This is broadly consistent with findings from de Renzio et al. (2011) and Andrews (2010) – statistical significance is somewhat improved in our analysis reflecting availability of more data, and we find a smaller impact of GDP on PFM quality than reported previously.

regression is lower given that to date only a limited number of countries have undertaken repeat PEFA assessments. Running a robustness check using the CPIA indicator for PFM quality (CPIA-13), the results are similar with regards to a significant and negative relationship with initial PFM quality (albeit somewhat smaller). In addition, a strong negative effect of the change in population becomes significant, suggesting that countries with faster population growth have done worse with regards to improving the quality of their PFM systems. Furthermore, the fact that resource dependence does not show a significant relationship with in the first difference analysis using PEFA data, and a weak and barely significant relationship using CPIA data, implies that being resource dependent does not prevent a country from progressing with PFM reforms relative to progress seen in other countries—which is encouraging. Still, due to data constraints there is greater uncertainty regarding this finding, and the analysis should be revisited as repeat PEFA assessments for additional countries and time periods become available.

Both the cross-sectional and first-difference findings suggest that to some extent, differences in levels and changes in PFM performance are associated with country characteristics; while only a few variables combine high statistical significance (at the 1 percent level or higher) and a tangible impact. To understand more about why PFM reforms progress in some countries but less so in others despite similar broad characteristics, such as having a similar income level and regime type, we will need to shift from large N analysis using standardized data to comparative case studies that allow analyzing specific factors, and the dynamic interaction between institutional development and stakeholders to explore whether these account for such differences.

For the third research question, a preliminary investigation is conducted into the relationship between PFM quality and key fiscal outcomes, including cost-effectiveness in the use of funds for public service delivery (technical or operational efficiency). Findings validate that better public financial management is associated with lower variances in overall budget execution rates as well as cross-sector variance (robust when controlling for GDP per capita), but not with lower deficits. Furthermore, we do not find a correlation between PFM quality and public service delivery outcomes in health or education relative to these sectors' fiscal allocations—that is, available data do not show a significant relationship with the operational efficiency of spending.[4] Given the data limitation discussed in the main sections, this part of the analysis needs to be interpreted with the greatest care.

The findings of this paper should be interpreted cautiously due to potential endogeneity and OVB. We do undertake various robustness tests in an effort to account for potential measurement error and OVB, including a variety of specifications and alternative sources of data. Regardless, the findings should not be interpreted as evidence of the direction of causality and more testing of alternative model specifications is needed.

This paper is the first part of a wider study on *Making reforms of PFM work: political economy aspects of buy-in and follow-through*, aimed at exploring the role played by political economy factors in enabling or hindering PFM reforms and their effective implementation. It is increasingly recognized that such factors play a crucial role in differentiating whether and to what extent PFM reforms succeed. In this first part, the focus is on empirical analysis of country

---

[4] These findings are consistent with Vlaicu et al. (2014) and World Bank (2013) which focus on the impact of a specific component of PFM systems—medium-term expenditure frameworks—on fiscal performance.

level data on PFM quality and associated country characteristics. Further work under the study will look at how political economy factors interact with key socio-economic and political drivers of PFM quality.

The paper proceeds as follows. In Section 2, we present our analytical framework for the paper, including a brief overview of relevant literature. Sections 3 and 4 describe the empirical models to be tested, our methodology and assumptions, and the data sources used in the analysis. In Section 5, we present the results of the econometric analysis. Section 6 follows with concluding remarks.

2. The state of knowledge on drivers for the quality of public financial management

2.1 Drivers of PFM performance

The relationship between country characteristics and PFM performance has begun to be explored by some recent analysis as relevant cross-country data on the latter has become available. This paper builds on these efforts. De Renzio et al. (2011) undertook an initial analysis using accumulating PEFA data as part of a major donor evaluation effort concerning the effectiveness of PFM reform support, based on a global sample of 90+ countries with available PEFA scores. The main relationships diagnosed as significant are GDP per capita, recent GDP growth, population size, natural resource dependency, state fragility, and the level of PFM support provided by donor agencies. Among these, the two stronger effects emphasized by de Renzio et al. are levels of GDP per capita (positive) and state fragility (negative). Population size is also a significant factor, while PFM donor support, though statistically significant, has only a small explanatory effect. The paper also rejects a number of hypotheses, most importantly trade openness, colonial heritage, regime type, and aid dependency, which are all shown not to have significant effects in the multivariate cross-country analysis.

Another related paper, by Andrews (2010) focuses on a narrower sample of 31 Sub-Saharan African countries. For this group, Andrews finds a strong relationship between recent growth rates (over the decade 1996 to 2006 – that is, up to the years prior to the global financial crisis) and average PEFA scores. With regards to income levels, he notes that for the full sample of 73 countries for which PEFA scores were available at the time, higher income countries (with per capita incomes above $4,000 per capita) are more likely to attain top PEFA scores than those with lower incomes. But, this relationship between better PEFA scores and higher per capita incomes did not hold within the SSA sub-sample of 31 countries. A second main finding is that fragile states within the Sub-Saharan African group were among those with the lowest PEFA scores. Andrews also notes, however, that some post-conflict countries such as Sierra Leone and earlier Mozambique have shown an ability to rapidly improve their PFM systems once stability was regained.

Furthermore, Andrews (2009) quantifies a number of country factors to analyze their effect on PFM change in Africa through a multivariate ordered logistic model.[5] He finds several country-

---

[5] Andrews derives findings on drivers of PFM quality change from cross-country comparisons. This paper, in contrast, looks at changes in PFM quality within countries over time.

level characteristics as significant including natural resource dependency (negative), the length of countries' commitment to PFM reforms (positive), and colonial heritage (Francophone = negative). Overall, he concludes that there are relatively strong effects of country context variables for PFM performance.[6] Finally, he indicates that HIPC played an important role in generating political commitment to PFM reforms in several SSA countries.

Our paper evolves the approaches used by the previous work in the following ways. First, we revisit the main relationships discussed with somewhat differing findings by the previous literature (Andrews (2010), Andrews (2009), de Renzio (2009), and de Renzio et al. (2011)) [7] using a growing pool of data. We explore some additional variables that have been suggested as relevant for driving PFM performance, in particular the level of domestic revenue, the experience of macro-economic and fiscal shocks, and the presence of programmatic parties (question 1). Second, we add an analysis of 'first differences', i.e. of comparing how the results of PEFA assessments have changed between different rounds, relative to changes in country characteristics (question 2). Third, because the effects of PFM improvements are a key part of the motivation for seeking such improvements, we add an investigation of this relationship – which has been identified as an important gap in the empirical knowledge as pointed out by Pretorius and Pretorius (2008) (question 3). The next sub-section describes the background to this component further.

2.2 Effects of PFM performance

PFM reforms have three broad objectives: (i) to improve aggregate fiscal discipline – budgets should be consistent with a realistic macroeconomic framework and a sustainable fiscal program, and brought in on target, (ii) to improve allocative efficiency (or strategic prioritization) – that is, an allocation of funds that is aligned with country priorities and helps maximize social welfare,[8] and (iii) to improve operational efficiency – requiring that resources are utilized efficiently and effectively towards to the purpose for which they have been allocated without waste, loss due to corruption and other forms of leakage (see among others Tommasi 1999).

Despite the fact that these three objectives are central to the PFM practice – and form part of the basis for the PEFA framework – the relationship between PFM performance and these

---

[6] "The strongest message from this section is simply that country characteristics matter a great deal in understanding what PFM system quality looks like."

[7] De Renzio et al. (2011, p.17) find that a doubling of income resulting in a 0.5 point increase on the 1-4 scale for PEFA scores. This would imply that GDP would account for much of the PFM quality differences between countries. However, this finding relies on an error in the computation of the effect of GDP differences. Correcting for this, their estimates imply a much smaller impact of GDP, closer to what this paper finds (see Section 5.1). Note that the question of whether PFM systems can be equally well performing across countries with very different income levels is often bracketed by the reform practice – for example, PEFA assessments have been developed as a standard gauge to be applied across different income levels.

[8] There is an implicit expectation that government programs are at least broadly aligned with the goal of improving social welfare. A political economy perspective reminds us that this assumption may not hold true in many situations; or that the alignment is at least limited, and in some countries more so than in others.

results has not been well researched empirically. On the one hand, there is a wealth of specific findings that weaknesses in PFM affect the ability of a government to deliver services to their citizens – as, for example, set out in many *Public Expenditure Reviews* (PERs) and *Public Expenditure Tracking Surveys* (PETS). But efforts to explore the cross-country experience on the link between PFM quality and reforms, and (improvements in) fiscal outcomes and service delivery have been scant. In their comprehensive review of published materials on PFM reforms, Pretorius and Pretorius (2008) note this as the first area where they see a significant gap and a need for greater investments in building a better evidence and knowledge base. This scarcity is primarily due to the fact that until very recently, cross-country measures of the quality of PFM systems were not readily available, as well as data limitations with regards to the allocation of funds and their actual usage – in particular with regards to the efficiency of spending.

The empirical work that has been done has primarily focused on the effects of improved public financial management on aggregate fiscal discipline – both because this is a widely shared concern, and because the relevant data on fiscal deficits and changes in debt have been more readily available (e.g. von Hagen and Harden 1996; Prakash and Cabezon 2008; Alesina and Perotti 1996). Most of the findings from these studies – focused on different groups of countries and time periods – confirm a relationship between better PFM systems and a more sustainable fiscal balance, albeit with various caveats and nuances.

Among the first studies to explore the relationship between aspects of PFM reforms and all three goals are Vlaicu et al. (2014) and World Bank (2013). These studies look in particular at the relationship between the adoption of an MTEF and the following variables: fiscal discipline (which is part of the primary rationale of an MTEF), sectoral expenditure volatility (as a proxy for allocative efficiency), and cost-effectiveness of spending in the health sector (as a proxy for operational efficiency). Findings suggest that fiscal discipline improves after the adoption of an MTEF, some improvements in reduced volatility of allocations to the health sector, but no clear effect on operational efficiency as proxied for in the analysis.

The preliminary assessment of the impact of PFM system quality on fiscal outcomes and public service delivery undertaken in this paper is important to better capture what expected effects can actually be observed. This includes a political economy aspect: asking decision-makers in developing countries to commit to PFM reforms, to remain committed and to pursue effective implementation is significantly more persuasive if such reforms are linked to tangible results within reasonable time periods. The empirical basis for developing better evidence has been improving with the increased availability of more and better data on the quality of PFM systems across countries and points in time, as well as on expected impacts. Nonetheless, there are still a number of important data gaps that will need to be closed in the coming year so as to enable more reliable and granular analysis.

3. Empirical Models

3.1 Macro-Level Country Characteristics and PFM Quality

As set out in section 2 above, we start the exploration by investigating the extent to which GDP per capita appears as a fundamental driver of PFM performance, as this has been identified as a

key driver in previous studies. Next, we explore whether other macro-level country characteristics are associated with PFM quality by building on the previous literature. Our interest is to explore whether these variables show a relationship with the variance in the quality of PFM systems (across countries and within countries over time). We hypothesize that overall, macro-level country characteristics are relatively weak or insignificant in explaining the level and change PFM quality across countries and over time, ceteris paribus. This is based on the anecdotal observation that PFM reforms appear to have succeeded in a substantial range of different country contexts.

The cross-sectional model, across countries, is estimated as follows:

(1a) $P_i = \propto + \beta_{1i}I_i + \beta_{2i}X_i + \varepsilon_i$

The first-differences model focusing on within-country changes over time is as follows:

(1b) $P_{it} = \alpha + \beta_{1it}I_{it} + \beta_{2it}X_{it} + \delta_i + \varepsilon_{it}$

where i indexes countries, P is PFM quality, measured by the most recent PEFA assessment (using an average of 21 selected PEFA sub-indicators – see Section 4.1), I is the log of real GDP per capita, X is a matrix of socio-economic and political macro-level variables, δ is fixed effects, and ε is the error term. These equations are estimated using ordinary least squares (OLS).

For the first model, we are able to use a dataset of 120 countries to study the cross-country relationship between PEFA scores and a variety of country characteristics in levels. The dependent variable is the most recent PEFA score available (ranging from 2008 to 2012 for each country), and the independent and control variables cover a five-year lagged average prior to the most recent PEFA score for each respective country.[9]

Second, we examine changes in PFM quality for almost 50 countries for which we have at least two PEFA scores using first differences. We estimate this second model by computing the average annual changes for our dependent and independent variables between the first and most recent PEFA measurement.[10] We also perform a robustness check using CPIA-13 as our dependent variable instead of PEFA. Looking at changes within-countries over time indicates whether macro variables associated with the differences in PFM quality between countries also matter for changes over time (question 2). In other words, we are asking if X improves within a country, should we expect P to improve, and if so, by how much?

The econometric strategy outlined here does not demonstrate a *causal* link between macro-level economic and population variables, sources of revenues, political and institutional factors,

---

[9] This differs from previous work, which calculated a five-year average from 2002 to 2006 for all countries, despite the fact that the year of the assessment was not the same for all of the countries being compared, and in fact the range was quite extensive from 2005 through 2010. We only include those countries that have data for at least three of the five years.

[10] Since only a limited number of countries completed multiple PEFA assessments until recently, this paper is the first to undertake such an econometric analysis. The analysis includes PEFA assessments up until 2012. Assessments completed in 2013 are not included, to ensure availability of data for other variables.

on the one hand, and PFM quality, on the other hand. While we do mitigate concerns about endogeneity and omitted variables by employing the first differences analysis, the conclusions should still be interpreted with this caveat in mind. Findings on statistical correlation are nevertheless valuable for answering the questions that motivate this paper. Most importantly, the relative strength of the associations indicate to what degree PFM performance is likely to be 'bounded' by the set of variables explored here.

3.2 PFM quality and outcomes

Next, the linkages between PFM performance and key outcome variables are examined (question 3 of the Introduction). We hypothesize that higher PFM quality is associated with better fiscal discipline (i.e., lower deficits to GDP as well as aggregate budget execution rates close to planned levels), ceteris paribus. We also expect that better PFM systems result in better allocative and operational efficiency, consistent with the literature outlined in Section 2.

We investigate these relationships using simple OLS cross-sectional regressions by estimating the following model:

$$(2) \ Y_i = \propto + \gamma_1 P_i + \ \gamma_2 I_i + \ \varepsilon_i$$

where i indexes countries, Y is the dependent variable of interest (measures of aggregate fiscal discipline, allocative efficiency and operational efficiency), P is PFM Quality, measured by the most recent PEFA assessment (an average of 21 selected PEFA sub-indicators – see Section 4.1), the control variable I is log GDP per capita, and ε is the error term.

Similar to the first part of this analysis, we start with a dataset of over 100 countries to study the cross-country relationship between PEFA scores and a variety of outcome variables.[11] We assume that the quality of PFM systems as measured by PEFA assessments measures the accumulated capabilities and efforts at system reforms. For each country, we use the earliest PEFA score available (ranging from 2005 to 2012). On the dependent side, as described in section 2 above, we use available proxies for the three broad categories of expected effects: (i) for aggregate fiscal discipline – the annual deficit to GDP ratio and the overall budget execution rate (PEFA indicator P-1), (ii) for allocative efficiency – the sectoral variance in budget execution (PEFA indicator P-2), (iii) and for operational efficiency – the cost effectiveness of health and education expenditures.[12]

---

[11] The number of observations available for the dependent variables ranges from 56 to 102; see Annex 1, table A1.2.

[12] These variables are imperfect proxies for the three dimensions of fiscal performance. For example, fiscal discipline is not directly proportional to deficits—fiscally responsible countries with lower fiscal liabilities (explicit and contingent) and interest rates or facing a temporary shortfall in aggregate demand may run larger deficits than their comparators. Similarly, in select cases a low score on PEFA indicator P-1 for the budget execution rate may reflect that the country frequently adjusts budget implementation when macro-economic conditions change. These may be signs of good fiscal discipline instead of profligacy. Also, the deviation from sectoral budget plans during budget implementation (PEFA Indicator P-2) may reflect inefficiencies in budget planning—although this is more often an indicator of skewed allocative decision making.. But even with imperfect proxies, we can start a process of empirically exploring these relationships with the most relevant empirical data available.

As outlined above, our strategy in testing this second relationship does not deal with concerns about endogeneity and omitted variables, and does not establish whether there is a causal link between PFM quality to fiscal performance. [13] This implies results merely indicate association and need to be interpreted with caution.

For both relationships – *Drivers of PFM performance* and *Effects of PFM performance* – we perform a number of robustness checks in order to test the validity of our results. For the first model (1a and b), we check for potential measurement error in our independent and dependent variables by using different sources to proxy for PFM quality, income, revenue, aid dependence, political stability, and regime type. For the second model, we use an alternative specification for education. Further checks include different lengths of lags (and no lags) and alternative aggregation techniques for PEFA, and we run our regressions by including and excluding middle-income countries. A summary of the main results of these robustness checks is presented in Annex 3.

4. Variable Descriptions and Data Sources

Our sample includes the results of national-level PEFA assessments for 125 low- and middle-income countries over the period 2005 (when the PEFA was initiated) to 2013. Several countries were excluded in both explorations as too many indicators in their PEFA assessments were not rated, or other data limitations.[14] This resulted in an initial dataset of 120 countries in model 1a, 42 countries in model 1b, and 102 countries in model 2 (see Annex 5 for the list of countries).

4.1 Measures of PFM Quality

To assess and compare the quality of PFM systems across countries and over time we considered two datasets including the World Bank's Country Policy and Institutional Assessment (CPIA) indicator 13, and the set of indicators developed under the PEFA initiative.

CPIA-13 is one of 16 criteria established by the World Bank as part of an annual diagnostic tool that aims to capture the quality of a country's policies and institutional arrangements. CPIA-13 measures the quality of budgetary and financial management on a six-point scale along three dimensions: (1) a comprehensive and credible budget, linked to policy priorities; (2) effective financial management systems to ensure that the budget is implemented in a controlled and predictable way; and (3) timely and accurate accounting and fiscal reporting, including audit.

---

[13] Issues of endogeneity and omitted variables are addressed in Vlaicu et al. (2014) to examine a similar relationship (between the introduction of medium-term expenditure frameworks and fiscal performance) through instrumentalization and panel techniques, focused on comparing fiscal performance before and after the introduction of medium-term expenditure frameworks. A comparable approach cannot be applied here due to the structure of the data which lacks a 'before-and-after' dimension.

[14] We excluded any countries missing more than 3 out of the 21 PEFA indicators used for our estimate of PFM quality including Lebanon, Macedonia, Nauru, and St. Pierre and Miquelon (See Annex 5 for the list of PEFA indicators).

The PEFA Measurement Framework identifies six dimensions of performance: (1) credibility of budget; (2) comprehensiveness and transparency; (3) policy-based budgeting; (4) predictability and control in budget execution; (5) accounting, recording and reporting; and (6) external scrutiny and audit. For these core dimensions, a set of 28 indicators, including 65 sub-indicators, are used to assess the performance across these six dimensions. The PEFA framework also includes a set of 3 indicators measuring donor performance not considered here.[15]

PEFA data are used in this paper as the principal indicator of PFM quality; however, we do use CPIA data to check the robustness of our results. We believe PEFA data are based on more comprehensive and detailed in-country assessments and therefore can be assumed to reflect countries' PFM performance more fully. They also provide more information than CPIA scores including the possibility of exploring the breakdown into the six dimensions.[16] After nearly a decade of implementation, PEFA (with 120 low and middle income countries) now offers country coverage comparable to the CPIA dataset (135 countries). The correlation between CPIA-13 and the aggregated PEFA score that we use is quite high at 0.764.

Figure 1. Aggregate PEFA Score by Dimension
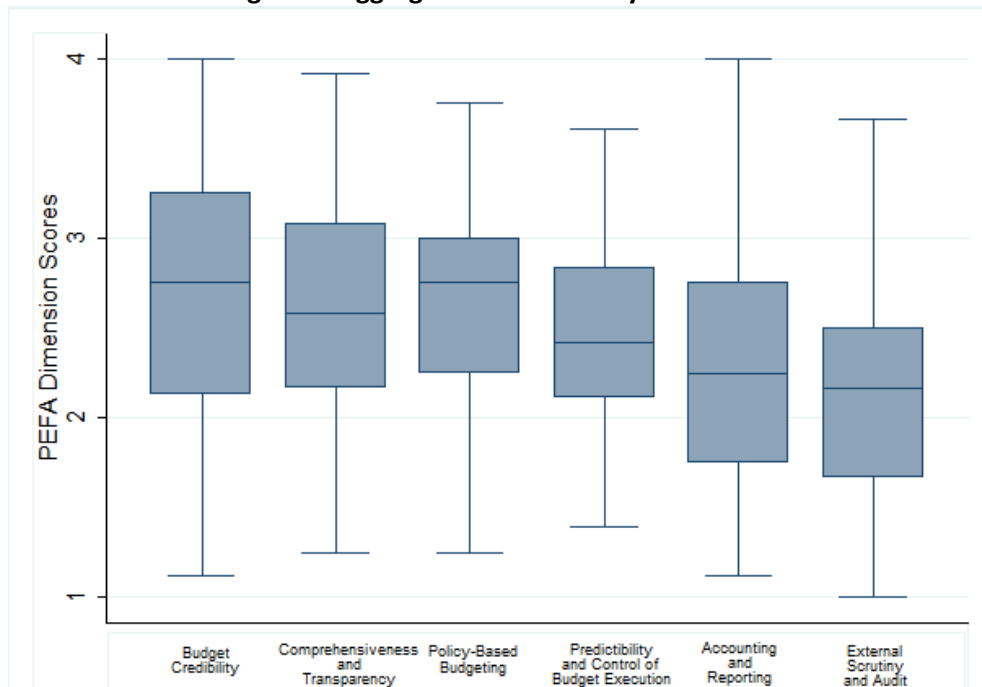
Figure 1 compares the distribution of PEFA scores for each of the six dimensions across the 120 countries in our sample (using the most recent assessment for each country). Among the six dimensions, the external audit and parliamentary oversight dimension has the lowest average

---

[15] See www.pefa.org for the full description of the PEFA Framework, and also selected methodological adjustments that have been made over time.

[16] It should also be noted that in recent years, CPIA-13 has been drawing on information contained in PEFA assessments. See http://www.worldbank.org/ida/IRAI/2011/CPIAcriteria2011final.pdf, p. 39 and PEFA Secretariat (2010).

score while budget credibility and policy-based budgeting have the highest scores.[17] For each of the dimensions, and in particular for the accounting and reporting dimension, the spread of country ratings is wide.

To aggregate the PEFA data into a single number for our analysis, we summarize those PEFA indicators that cover the quality of PFM systems on the expenditure side. We therefore exclude PI-1 through P1-4, which measure PFM outcomes, indicators PI-13 to PI-15, which cover transparency and effectiveness of tax administration, and D1 to D3, which are donor-related indicators. This leaves 21 indicators (PI-5 to PI-12 and PI-16 to PI-28) for our analysis (see Annex 3 for a complete list). We then converted the alphabetic indicator ratings included in the PEFA assessments into numerical values, with higher scores denoting better performance (i.e. A = 4 to D = 1), and then calculated a simple average of the selected 21 summary indicator scores, assuming equal weights for each indicator – that is, all indicators have equal importance – in order to arrive at one measure of overall PFM performance.[18]

While there are validity concerns about converting ordinary PEFA letter ratings into numbers and aggregating the individual indicators, we believe that this is the best approach for the aims of the analysis presented here.[19] This approach follows De Renzio et al. (2011) with some variation.[20] Alternatively, a multivariate ordered logistic model could be used (cf. Andrews 2009), but results would be more difficult to interpret. Regardless, we undertake sensitivity analysis to highlight the extent to which our findings are robust under alternative ways of aggregating PEFA scores, and found no significant differences in results.

A brief assessment of how our measure of PFM performance is associated with overall governance effectiveness – as measured by the World Governance Indicators – is presented in Figure 2.[21] Given that PFM systems are one part of the overall concept of government effectiveness, we would expect a close relationship between these two variables. Furthermore, given that other aspects of government effectiveness may face even greater reform challenges,

---

[17] The boxes contain the middle 50% of the data (bordered at the 25th and 75th percentiles of the PEFA score with the line representing the median value of the data). The ends of the vertical lines (whiskers) indicate the minimum and maximum data values to a maximum of 1.5 times the inter-quartile range values (over 97% of the data in our case). Outliers beyond this range are excluded from the graph.

[18] In order to avoid misleading results, we excluded countries that were missing more than 3 indicators or "no scores" due to lack of adequate information or because they were deemed to be inapplicable in the PEFA assessment.

[19] The PEFA Secretariat has issued a report summarizing various methods used while taking an agnostic stance as to which are valid or superior. PEFA Secretariat (2009), *Issues in Comparison and Aggregation of PEFA Assessment Results Over Time and Across Countries.*

[20] The present approach differs slightly from de Renzio et al. (2011) by averaging the 21 indicators instead of using an average of the 64 sub-indicator scores. The latter approach by de Renzio et al. puts more weight on those indicators with multiple sub-indicators compared with dimensions with few or no sub-indicators.

[21] An important caveat is the fact that government effectiveness is a composite concept, and that good measurement of the concept remains a challenge that has not yet been fully met by existing indicators. The WGI GE indicator is a composite of several other indicators, and draws among others on indicators measuring the quality of PFM, but not on PEFA data itself. We rescale the government effectiveness variable (percentile rank) to 0 through 10 for easier comparison with other variables in this paper. The higher levels denote greater government effectiveness.

and that over the past decade or so attention to PFM has been particularly intense, one might expect measures of PFM performance to exceed those of overall government effectiveness.

We find a correlation of 0.4607 and a regression coefficient of 0.1166 (at 99 percent level) in a simple bivariate regression. We also find a number of ECA countries above the line – suggesting that they may have reformed their PFM systems to a greater extent than other aspects of overall government effectiveness. There is also a group of outliers towards the lower right-hand quadrant (high government effectiveness but low PFM), which are mostly small island states (Antigua, St. Vincent and the Grenadines, and Saint Lucia). This may be an indication that such countries, where conducting the business of government may be less complex than in larger countries, may not need to meet strong PFM standards to achieve overall government effectiveness. Excluding these three outliners, the correlation rises to 0.5506 and the regression coefficient to 0.143 (also at 99 percent level).

Figure 2: PFM Quality and Government Effectiveness
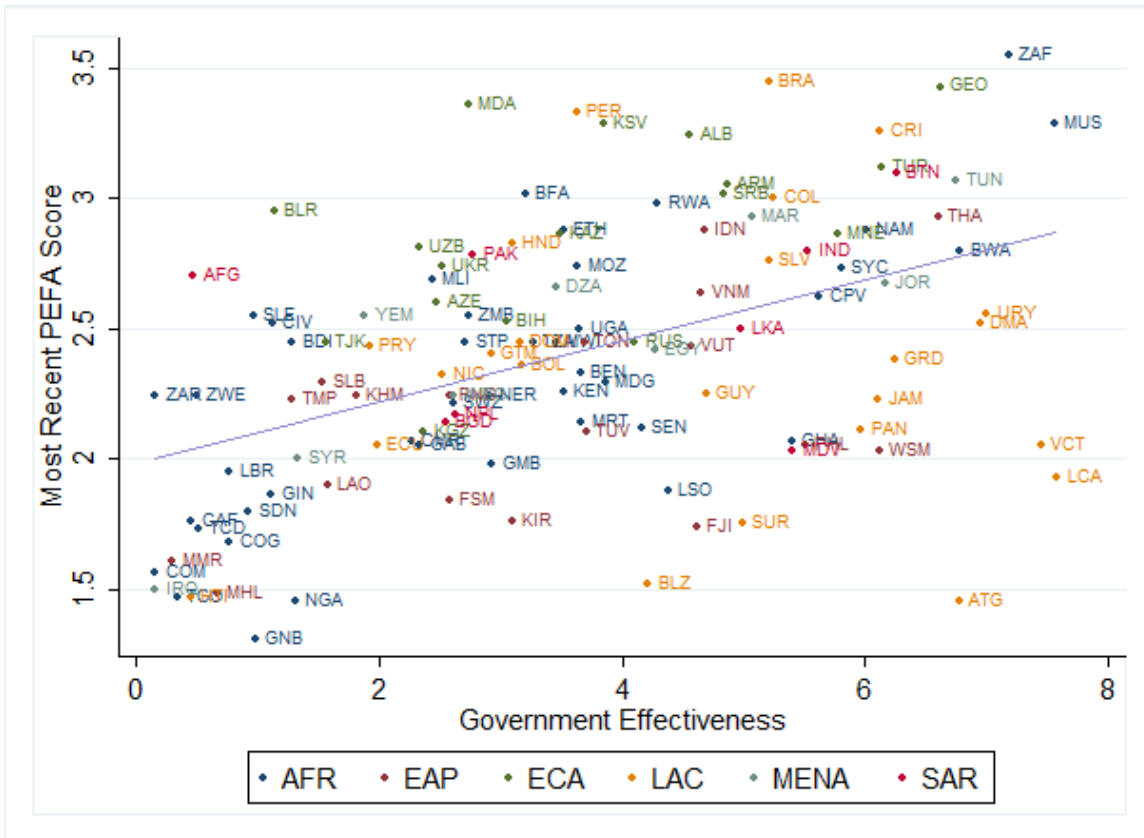
4.2 Country Characteristics Associated with PFM Quality

We explore the potential association between ten variables and PFM reforms (the main variables are set out in table 1 below). These fall into four broad categories: (i) the economic level and trend (GDP per capita, growth, and resource dependence, and shocks); (ii) population

and being a small island state; (iii) the sources of funding (tax revenue and aid dependence); and (iv) macro-political characteristics: stability, regime type, and the presence of programmatic parties. These overlap with variables explored by de Renzio et al. 2011 and with Andrews 2009. We do not test for trade openness, technological diffusion, levels of adult education, and for administrative heritage (being a former French or British colony), which did not show consistent statistical significance in previous analyses.[22] We re-test for aid dependency despite the fact that de Renzio et al. do not find a clear relationship, given rather strong expectations about an association; and similarly for regime type, where we also explore a possible relationship with budget accountability specifically. Variables that we newly explore are in particular the level of tax revenue, and the presence of programmatic parties, both of which are either based on strong assumptions of relevance (tax) or have been shown to be significant for closely related areas (programmatic parties for overall public sector quality). Furthermore, we explore whether shocks (fiscal or macro-economic) seem to motivate PFM reforms. Given the increase in the sample size, our analysis furthermore offers an opportunity to revisit the robustness of the findings of earlier work.

Income per capita is measured as the log of real GDP per capita in purchasing power parity (PPP) terms at constant 2000 prices. As discussed, income is associated with a wide range of variables which would enable better PFM systems, including financial, human and technical resources, and may also be associated with demand for better fiscal performance and public service delivery (see the below discussion on the revenue variable).

Growth is measured as the annual percentage growth rate of GDP per capita (five-year average, lagged by one year prior to the most recent PEFA score). Higher rates of recent growth are expected to facilitate institutional improvements through its impact on resource availability and possibly growing expectations of what government ought to achieve. We also check whether the experience of growth shock in recent years prior to a PEFA assessment show an association.

Additionally, we check whether the experience of growth or of fiscal shocks has a significant impact (within four to five year lags). The expectation in this regard is that such shocks could stimulate efforts at strengthening PFM systems, as a way to increase institutional resilience and the ability to use tighter resources better.

Comparison of PEFA scores with government effectiveness suggests that *population* size relates to PFM quality. De Renzio et al. (2011) find a positive and significant relationship between population size and PFM quality. Furthermore, countries with particularly small populations, *Small Island Developing States (SIDS)* may lack resources (financial and human) and have less need for advanced PFM systems (see Section 4.1). Similarly, larger states may find the cost of centralized PFM systems to be low in relation to their revenue base and the return to investment, which impacts the performance of a large budget, to be high.

---

[22] See de Renzio et al. (2011), 16-18. Trade openness and French colonial heritage showed a weak significance in their simple model, but not the comprehensive model. Andrews (2009) finds a significance of Francophone heritage (negative) for the smaller sample of African countries included in his analysis. Andrews includes 'length of reform commitment' in his analysis, and finds it to be significant, measured as when a country first formalized a PRSP; but we think that this measure may not be a sufficiently good proxy when including countries from different regions.

The *resource* dependency variable is a dummy variable capturing those countries classified as 'resource rich' if more than 20 percent of total revenues is derived from natural resources. We determined these classifications based on data and previous work of the International Monetary Fund.[23] Resource dependence is assumed to have negative impacts through a variety of channels, including reduced accountability between citizens and state elites, greater incentives for intransparency in the management of public funds, and the presence of windfall revenue as well as revenue volatility which might negatively affect budget planning and execution (de Renzio et al. 2009, Daban Sanchez and Helis 2013).

The *tax revenue* variable measures the level of tax revenue as a percentage of GDP. This variable uses a database compiled by the International Monetary Fund's Fiscal Affairs Department (drawing on GFS, WEO, and OECD data). Higher levels of domestic revenue would provide resources to governments to invest inter alia in better PFM systems. Moreover, as citizens pay more taxes, they expect governments to use these funds well and in ways that result in improved and expanded services (Moore 2004 and Prichard and Leonard 2010). This should incentivize governments to invest in PFM systems that facilitate delivery of results and enable greater accountability. The connection rests on the assumption that the tax revenue derives from citizens rather than from natural resource wealth, since the latter may undermine rather than strengthen accountability relationships (see Andrews 2010 and Auty 2000). The main variable we use includes mainly non-resource related forms of revenue.[24]

The *aid* variable measures the level of official development assistance as a fraction of GNI. The source is the World Bank's World Development Indicators, which contains, among other things, all of the official loans and grants received by developing countries from multilateral or bilateral sources.[25] For this variable, there are two related but somewhat contradictory expectations. One is that aid dependent countries are more likely to undertake public sector (including PFM) reforms (Therkildsen 2000 and 2001; see also Fialho Lopes and Fritz 2012), or that there is possibly even a reverse causation with donors investing more in countries that show greater effort and success with reforming their PFM systems. A contrasting expectation is that aid dependent countries invest less in such reforms relative to similar countries that rely more on domestic revenues, due to stronger accountability relationships in the latter and negative effects of aid dependence on the coherence of the public sector (Brautigam and Knack 2004 and Moss et al. 2006). Thus, the exploration here is whether any clear cross-country pattern, either positive or negative, is statistically significant. To test for potential measurement error of ODA and to check our results, we substitute aid with both Country Programmable Assistance (which excludes volatile aid such as debt relief and emergency relief as well as donor overhead cost which is spent outside of the recipient country) and Technical Cooperation (see Table A3.1, columns 4 and 5). Finally, we check for inclusion and completion of the HIPC process, to explore whether this group of countries differs significantly in terms of PFM performance from others.

A first macro-political driver that we test is *political stability.* Political stability is widely considered a necessary ingredient for developing and improving institutions; and as mentioned in section 2, earlier work by de Renzio et al. (2011) has confirmed the opposite, i.e. fragility, to

---

[23] See Baunsgaard, Villafuerte, Poplawski-Riberio, and Richmond (2012), "Fiscal Frameworks for Resource Rich Developing Countries," International Monetary Fund.

[24] Royalties are excluded; while company income tax paid by resource-related companies are included.

[25] We cross-checked and verified the consistency of the data with the OECD database.

be significantly and negatively related to PFM performance. De Renzio et al. use the presence of UN peace-keepers since 1995 to measure fragility. We start by using a continuous measure, the rating for political stability by the Worldwide Governance Indicators (Kaufman, et. al). We rescaled the percentile rank data to the range 0 to 10, with higher levels reflecting greater political stability (common units with other variables make comparisons of the slopes across different institutional variables easier to interpret). Political stability and absence of violence measures "perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means, including politically-motivated violence and terrorism." We test robustness with two alternative measures: one, 'fragility' by looking at which countries have been included in the World Bank's list of fragile states between 2004 and 2012. The variable measures the number of years a country was classified as 'fragile' in the five-year period before the latest PEFA assessment (lagged by one year).[26] Second, we use Foreign Policy' Failed States Index (Annex 3). This is a continuous variable based on a country's score.

The next variable of interest, *regime*, measures the nature of the political regime. As the main data source, we use the Freedom House index, which annually assesses each country on their political rights and civil liberties through expert assessments. We use a second data source, Polity IV, as a robustness check. To compute the variable, we aggregated Freedom House's two sub-indices, political freedom and civil liberties, which range from 1 to 7 (lower numbers indicate higher levels of freedom). We then rescaled and inverted this variable to the range 0 to 10, so the higher levels denote more political freedom. A more democratic regime could be expected to be associated with more effective accountability and through this mechanism with better PFM systems (de Renzio 2009; Lake and Baum 2001). However, from existing work, we know that the regime type as such is not very likely to show a statistically significant relationship with the performance of PFM systems. We test in addition, whether the degree of political competition is associated with performance on the accountability sub-dimension of PFM, i.e. the quality of external audit and parliamentary follow-up. As shown in section 4.1, external audit and follow up is the weakest of the six PEFA dimensions across all countries. We are interested to see whether the presence of a more democratic regime alters the performance on this dimension; given that in principle it should enable more independent audit offices as well as a stronger role of the legislature.

A third political variable considered here is the presence of *programmatic political parties*. The presence of programmatic parties signals that interest aggregation in a country follows a relatively stable pattern, and that interests are aggregated in ways that transcends individual personalities and instead is centered around 'programs' or policy stances. The measure cuts across the democracy-autocracy distinction; and it has shown to be significantly related to general public sector performance in existing research (Cruz and Keefer 2010), but it has not been included in previous research on PFM specifically. The variable used here ('parties') is constructed in the same manner as Cruz and Keefer (2010) using variables from the Database of Political Institutions (2012). It measures the share of the largest three government parties and the largest opposition parties that are 'programmatic' (right, left, or center) in their orientation. In other words, if a country has three out of four of these parties coded as left, right, or center,

---

[26] The World Bank's classification is in turn based on the overall CPIA ratings, with countries having a CPIA of 3.2 or below being classified as fragile. These countries were categorized as 'LICUS' in 2004 and 2005, 'Fragile States' in 2006 through 2009, and 'Fragile Situations' since 2010.

and only one that is either not discernible or inexistent, then it carries a value of 0.75 for that year.

Table 1 below summarizes these correlates or right-hand-side variables that are explored in relation to their effect on the quality of PFM systems. For our cross-country analysis (to answer the first question of the Introduction), we use a five-year average of the same variables, lagged by 1 year prior to the most recent PEFA score for each respective country.[27] In the first-differences analysis (to answer the Introduction's second question), we take the average annual percent change between the earliest and latest observation of the variables of interest (according to the dates of PEFA assessments).

Table 1. Summary of Potential Factors Related to PFM Quality			
Variable	**Theory/Hypothesis**	**Data Source**	**Variable Description**
Economic level, trends, and resource dependency			
Income Per Capita	Is widely associated with institutional quality, of which PFM is a component (Bluhm and Szirmai 2012, Acemoglu 2008). In the case of PFM, likely channels of interaction include resources and citizen demand for results which in turn requires PFM quality.	World Bank WDI	Annual GDP per capita, PPP in natural log (five year average, lagged by one year)
Growth Rate	Higher rates of recent growth are expected to facilitate institutional improvement	World Bank WDI	Annual percentage growth rate of GDP (five year average, lagged by one year)
Resource Dependency	Countries in which natural resources are more dominant in the economy are expected to have worse PFM systems (Andrews 2010; Auty 2000)	Baunsgaard, et. al. (2012)	Dummy variable for countries that rely heavily on oil/mineral revenue (>20% of total revenue)
Population characteristics			
Population	High returns-to-scale of investments in PFM associated with population size	World Bank WDI	Natural log of total population (five year average, lagged by one year)
Small Island Developing States	Similar to population, but focusing on the specific group of countries where government effectiveness tends to be high relative to PFM quality (see Section 4.1)	United Nations[28]	Dummy variables for countries that are classified as small island developing states
Sources of revenue			
Tax revenue	Greater tax based revenue is expected to lead to better PFM through stronger citizen demand for better services and greater accountability (Moore 2004, Prichard and Leonard 2010)	IMF FAD Tax Policy Database	Tax revenue as a percentage of GDP
Aid	Countries receiving more aid may invest more in improving PFM systems in order to continue or access more aid (Therkildsen 2000); but conversely, there are also	World Bank WDI	Official Development Assistance (ODA) as a share of gross national income (%)

---

[27] This differs from previous work by de Renzio et al. 2011 which calculates the average over 2002-06 for all countries, while the year of the first assessment is not the same for all of the countries being compared (this runs from 2005 through 2010. The model proposed here ensures the same lag structure between PFM quality and country characteristics across countries, with country characteristics (which are slow to change) leading PFM quality (which may change more quickly).
[28] See United Nations SIDS listing here: (http://www.un.org/special-rep/ohrlls/sid/list.htm)

	expected negative effects of higher levels of aid dependency (Brautigam and Knack 2004)		
Political/institutional variables			
Political stability	Greater political stability may facilitate sustained improvements in institutions, including PFM systems	Worldwide Governance Indicators (WGI)	Political stability indicator (rescaled: 1=least stable, 10=most stable)
Regime	In more democratic regimes, citizens are expected to have more opportunities to hold governments to account, creating an incentive for politicians to seek PFM improvements (de Renzio 2009)	Freedom House	Sum of Civil Liberties and Political Rights Index (converted to: 1=least democratic, 10=most democratic)
Programmat ic Parties	Programmatic parties are expected to be able to make credible commitments about improving public service delivery; and as part of this should have an incentive and capabilities to improve PFM systems.	DPI, Beck et. al 2001 and Cruz & Keefer, 2010	Fraction of parties in a country that are programmatic (either left, right, or center)

Table A1.1 in the Annex summarizes the number of observations available for each of the presumed 'driving' variables for PFM performance, the minimum and maximum range actually observed for the variable as specified, and the means and standard deviations. As previously noted in section 3.1, we use a dataset of over 100 countries to study the cross-country relationship between PEFA scores and a variety of country characteristics. In the first-differences model, we look at within country variation over time for about 40 countries, given the much more limited number of countries for which at least two PEFA assessments are available.

Furthermore, table A2.1 shows the correlation statistics for these variables. Income is related to a number of the other variables used, including a negative relationship to the level of aid to GNI, and positively to tax revenue, the presence of programmatic parties, and political stability. There is also some correlation among the political variables as might be expected, in particular a positive relationship between regime type and the presence of programmatic parties.

4.3 Effects of PFM Quality

The second relationship we examine is that between PFM quality and expected outcomes. As noted in section 2, this relationship is crucial from the perspective of the overall rationale for PFM reforms, but empirical work on this has been scant to date. Table 2 presents a summary of the left-hand-side variables that proxy fiscal performance along its three dimensions. The summary statistics of these variables are presented in Annex 1.

Following Vlaicu et al. (2014), our first outcome variable is the overall fiscal balance of a country, or *deficit,* as a proxy for aggregate fiscal discipline. We use the general government primary net lending/borrowing as a percent of GDP according to the IMF's World Economic Outlook database.[29] The variable is calculated as the three-year forward average beginning the year of

---

[29] See: http://www.imf.org/external/pubs/ft/weo/2013/01/weodata/weoselgr.aspx.

the country's first PEFA score.[30] We expect a positive correlation between the aggregate PEFA score and the overall fiscal balance.

In addition, PEFA indicator *PI-1* is used as a supplementary measure of a government's ability to maintain aggregate discipline.[31] This is based on the logic that budget credibility, i.e. keeping outturns close to budget plans, is a pre-condition for continuous fiscal discipline over multiple years. PI-1 is derived by comparing actual total expenditure to the originally budgeted total expenditure, but excluding debt service payments and donor-funded projects. That is, it measures whether governments are able to plan aggregate expenditures ex ante and keep to the broad parameter during execution. According to the PEFA methodology, countries with actual expenditures deviations of less than 5 percent of budgeted expenditures in the last two or three years receive a score of "A" or "4." On the other end, countries with deviations between actual and budgeted expenditures greater than 15 percent in two or three of the last three fiscal years receive a D or "1."

PEFA indicator *PI-2* indicator assesses expenditure out-turns against the original budget at a sub-aggregate level – broken down by main functions or programs where available, or else by main government spending agencies.[32] It is here used as a proxy for allocative efficiency. An important function of the PFM system is to ensure that the government's priorities, which have been identified and allocated through the planning and budgeting process, are actually funded as planned during budget implementation. Ensuring that ex ante functional or sectoral budget allocations are credible and implemented close to plans is therefore a pre-condition for allocative efficiency to be realized.

Indicator P-2 is calculated using two dimensions. First, it measures the extent of reallocations between budget heads during execution during the last three years (excluding contingency items). Second, it takes into account the average amount of expenditure actually charged to the contingency vote over the last three years. According to PEFA, countries receive an "A" if: (i) variances in expenditure composition of less than 5 percent in at least two of the last three years; and (ii) actual expenditure charged to the contingency account of less than 3 percent of the original budget on average. Countries receive a "D" score with: (i) variance in expenditure composition exceeding 15 percent in at least two of the last three years; and (ii) actual expenditure charged to the contingency vote of more than 10 percent of the original budget on average over the last three years.[33]

Table 2. Summary of Fiscal Outcome Variables

---

[30] We selected to use a three-year average instead of a five-year average due to the timing of the PEFA assessments and the 2008/09 international financial crisis.

[31] While a PFM system does not determine a country's fiscal balance, and a government could decide to have a more expansionary fiscal stance even with a well-functioning PFM system; an efficient PFM system should enable the government to manage outcomes in alignment with its budget policy intentions.

[32] See PEFA framework: http://www.pefa.org/en/content/pefa-framework-material-1.

[33] The methodology for calculating P2 was changed in early 2011, through the addition of a further dimension (size of the contingency relative to the overall original budget). For consistency, we calculate all assessments using the initial P2 methodology only (i.e. dimension i for P2 for PEFA assessments done since early 2011). There is no substantial difference in the results.

Variable	Hypothesis	Data Source	Variable Description
Deficit	Improved PFM systems should enable better aggregate fiscal discipline	IMF WEO Database	General government primary net lending/borrowing as a percent of GDP
Overall budget credibility (PI-1)	Improved PFM systems should lead to better overall budget credibility – which is in turn a pre-condition for aggregate fiscal discipline	PEFA, indicator P-1	Aggregate expenditure out-turn compared to original approved budget
Inter-sectoral budget credibility (PI-2)	Improved PFM systems should lead to greater credibility of sectoral allocations – as a proxy for allocative efficiency	PEFA, indicator P-2	Composition of expenditure out-turn compared to original approved budget
Health Efficiency	Improved PFM systems should support a greater achievement of service delivery outputs per unit spent (operational efficiency)	Vlaicu et al. (2014)	Operational efficiency of government health expenditure, with output measured as life expectancy
Education Efficiency	Improved PFM systems should support a greater achievement of service delivery outputs per unit spent (operational efficiency)	Authors calculation based on UNESCO data	Operational efficiency of government expenditure, with output measured as primary education completion rates

We hypothesize that PFM quality, measured by the aggregate PEFA score (PI-5 through PI-28, excluding PI-13 through PI-15), is positively associated with budget credibility (PI-1 and PI-2) on average, ceteris paribus. In other words, the better and more effective a country's PFM system, the more likely total actual expenditures will reflect planned expenditures (i.e. the variance between total actual and budgeted expenditures will be small). We also hypothesize that PI-2 will have a positive relationship with PFM quality. If policy planning and the budgetary process are well integrated, and if budget execution follows established procedures, then the composition of expenditures should align closely with the original budgets.

The last two variables we include in the analysis of PFM on expected outcomes are the technical efficiency of health expenditure *(health)* and of education expenditure (*education*), i.e. the cost-effective delivery of key public services. We hypothesize a positive relationship between PFM quality and these two variables. Better PFM systems should support cost-effective service delivery through several mechanisms: through effective up-front planning of where funds are needed and how much, making funds promptly available when and where they are needed, by reducing leakage – such as ensuring that ghost workers are removed from the payroll – and by ensuring or incentivizing efficient use of funds – for example by ensuring that value-for-money is obtained in procurement.

The outcome variables are calculated by relating key outputs for each sector to government per capita spending, and estimating the technical efficiency scores from a parametric Stochastic Frontier (SF) model, using maximum likelihood.[34] In the case of health, we used a dataset

---

[34] SF analysis was originally modeled by Farrell (1957) and elaborated by Greene (2005).

developed by Vlaicu et al. (2014) which used life expectancy at birth as the output and government health spending per capita (PPP) as the input (along with covariates population density and OECD membership, and year fixed effects). Similarly for education, we computed the efficiency scores in the education sector using the SF approach modeled by Belotti et al (2012). The outcome of interest is primary school completion rates and the input is government spending on education per capita (PPP), both measured by UNESCO Institute for Statistics.[35] Population density, from the World Bank Development Indicators, was used as a covariate.

It is important to note that the detailed causal chains between PFM quality and operational efficiency of service delivery cannot be mapped comprehensively. Since there are a number of other factors involved (including how health sector staff is trained and managed, the disease burden, etc.), the effects of PFM performance would have to be quite strong in order to appear as statistically significant in cross-country analysis.

5. Econometric Analysis

Following the research questions formulated at the outset, this section presents the econometric evidence on: (1) which macro-level country characteristics are associated (or not) with better PFM quality– using cross-section regressions (Section 5.1); (2) how changes in PFM quality over time relate to country characteristics using first-differences regressions (Section 5.2); and (3) observable effects of PFM quality on expected outcomes using cross-section regressions (Section 5.3). Additional quantitative analyses and robustness checks are shown in Annex 2.

5.1 Macro-Level Country Characteristics and PFM Quality: Cross-Section

To explore the relationship between macro-level characteristics and PFM quality, we first turn to the relationship between PFM quality and GDP per capita – which has been identified as a key variable associated with institutions in general and PFM quality in particular (see Section 2). The estimate of income on PFM quality using a simple bi-variate regression is significant at the 99 percent level with a coefficient of 0.17 and a standard error of 0.05 (Figure 3). While there is a statistically significant relationship, the magnitude of the relationship is rather small. This implies that changes in income levels have only a limited impact on PFM quality. For example, a doubling of GDP per capita is associated with an increase of 0.117 points on the (aggregate) PEFA scale of 1 to 4.[36] Comparing a richer country like Peru with a per capita income of nearly $8000 and a poorer country such as Bangladesh with a per capita income of around $1500, the

---

[35] For additional robustness, we also formulate the education efficiency variable using adult literacy rates.
[36] The expected change in Y associated with a 200% increase in X can be calculated as ln(2) x coefficient. Ln(2) = 0.693, so 0.693 x 0.169 works out as 0.117.

model would predict the difference in PEFA scores to be 0.27 points.[37] In other words, a 5-fold higher income is associated with only a 9 percent difference on the PEFA scale.[38]

Figure 3 below presents the bi-variate relationship between PEFA scores and GDP per capita across low and middle income countries. The figure reflects the positive and statistically significant relationship, but also there is large heterogeneity in the relationship with many countries far from the line. For example, both Belize and Peru, two countries in the LAC region, shared similar average levels of GDP per capita between 2004 and 2008; however, Peru's performance in PFM in 2009 is more than three standard deviations above that of Belize, which scores similar to both Iraq and Haiti. This also helps to clarify some of the seemingly contradictory findings between de Renzio et al (2011) that levels of GDP per capita make a considerable difference, and those by Andrews (2010) that among Sub-Saharan African countries income levels do not matter.

Figure 3. PFM Quality and Income
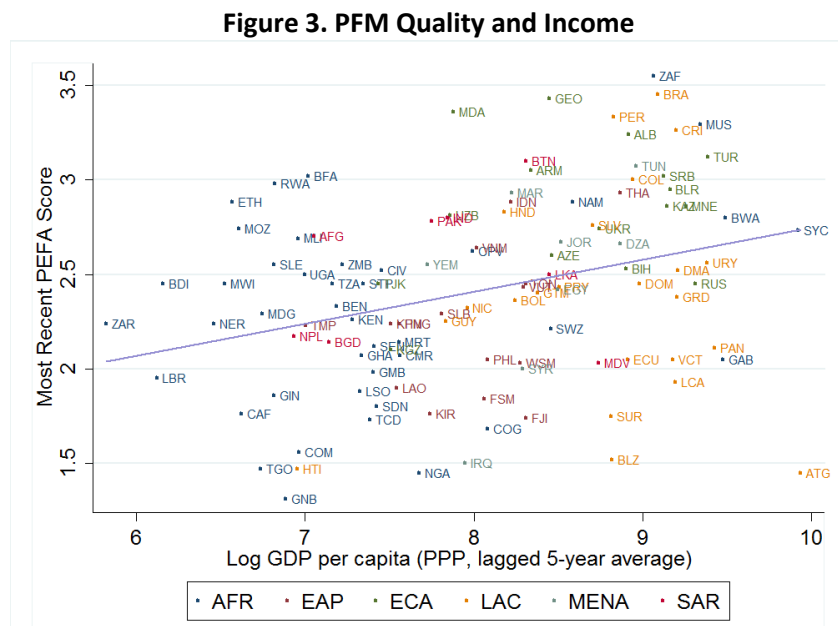
The three graphs in Figure 4 below further emphasize the wide variation in performance across countries by presenting the same bi-variate relationship disaggregated by income group. The first graph presents low-income countries and the latter two show lower-middle income and upper-middle income countries, respectively, with the lines representing the average PEFA score for each income group. As demonstrated, within income groups, heterogeneity is large. For

---

[37] Peru and Bangladesh have an average GDP per capita over five years (prior to their most recent PEFA assessments) of $7949.57 and $1568.44, respectively. To work out the expected difference in PEFA score associated with a 5-fold higher in income: Ln(5) x 0.169 = 0.272. This is equivalent to a 9 percent difference (0.272 divided by 3) on the 1 to 4 PEFA scale.

[38] Similarly, for a country with a starting income level of $750 per capita, and a 1.5 (i.e., score of D+) average PEFA score, the country would expect its score to increase to 2.0 (C score) when GDP per capita has increased tenfold to $7,500, ceteris paribus.

low-income countries, PFM quality has ranged from 1.31 for Guinea-Bissau to 3.02 for Burkina Faso, and for upper-middle income countries, PFM quality has varied from 1.45 for Antigua and Barbuda to 3.55 for South Africa.

Figure 4. PFM Quality and GDP per capita
By Income Group
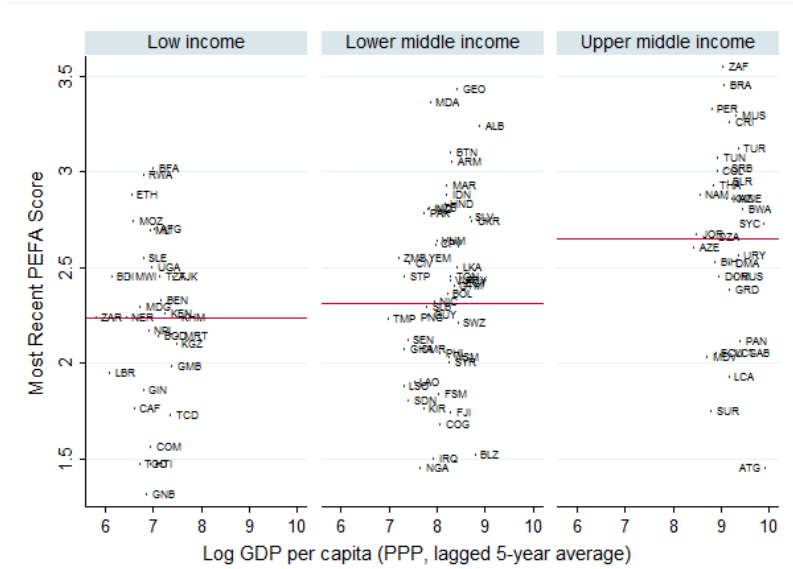
Figure 5 disaggregates this exploration further for the six sub-dimensions of the PEFA framework (like Figure 1 above, but divided by income group). External oversight remains the weakest sub-dimension in all income groups. It is clear that predictability and control in budget execution (dimension 4) as well as comprehensives and transparency (dimension 2) are higher among higher income countries, while policy based budgeting (dimension 3) is very similar across all levels of income. It should be noted that the difference between the median scores across dimensions is not statistically significant.

Figure 5. Aggregate PEFA Scores by Dimension
by Income Group
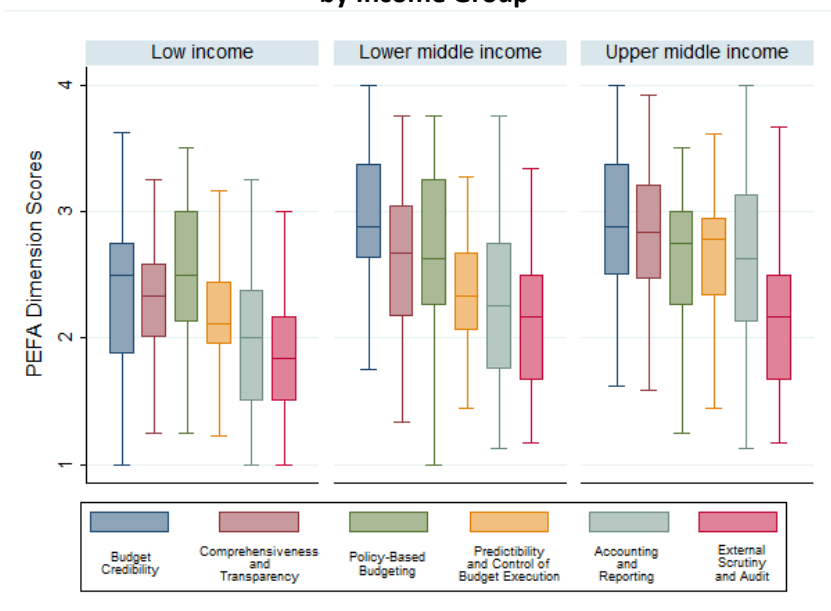
The regression's fit is improved considerably when we add other country characteristics discussed in section 4.2, as shown in Table 3. In column (1), we find that economic growth, population size, being a small island developing state, and resource dependency are all statistically significant in addition to GDP per capita, albeit at different confidence levels. Together, these country variables account for 40 percent of the variance in PFM quality (against 9 percent when only considering GDP per capita).[39] To provide an example, a country that is resource-dependent will, on average, have a PEFA score that is 1/3 point lower than non-resource countries as measured on the 1-4 PEFA scale (or almost ½ point lower on the CPIA 1-6 scale, see Annex 2).

We next add the additional variables related to sources of funding, tax revenue and aid dependency, and the three macro-political characteristics – political stability, regime type, and the presence of programmatic parties. Columns (2) through (6) confirm the findings in column (1) that the relationship between PFM quality and the first set of characteristics – i.e. the level of income, growth rates, population size, resource dependency as well as being a small island state remains significant. For the further variables tested, the results confirm several findings by earlier research, but also introduce several important differences and additions. The weak coefficient of tax revenue confirms assumptions found in the literature as discussed in section 4.2, but barely so. Aid is not associated with PFM quality, which is contrary to our expectations but in line with earlier findings.[40] We checked in addition for a possible association of PFM with

---

[39] As a robustness check, we test the relationship between PFM quality and income using GDP per capita constant 2005 $ rather than PPP and confirm our results. See Annex 3.

[40] This finding does not change for different measures of aid, including Country Programmable Aid (which excludes volatile aid such as debt relief and emergency relief as well as donor overhead cost which is spent outside of the recipient country) and aid in the form of Technical Cooperation (see Table A3.1,

prior growth and fiscal shocks (Table A3.1, columns 8-11), and with countries included in the HIPC initiative. We find no association with fiscal shocks, and the findings for growth shocks differ substantially depending on the time period used (four year versus five year lags). We suggest therefore that this issue should be revisited. We find a significant and strong negative association with the HIPC category; however, in this regard, the direction of causation may rather be the reverse, i.e. having poor PFM systems may have contributed to a country becoming highly indebted. The possible more specific effects of the HIPC process on improvements in PFM systems – typically an important condition for passing the decision point – are not tested for here.

With regards to the three political variables, we find a significant positive impact of political stability, a very weak relationship (in magnitude and significance) of the regime type, and for the presence of programmatic parties a relationship with weak statistical significance, but potentially substantial impact. The positive impact of stability is consistent with expectations and in line with earlier findings by de Renzio et al. (2011) (using a different measure). As noted in section 4.2, we check the influence of stability using two alternative specifications: the rating by the Failed States Index (Foreign Policy) and inclusion in the World Bank's list of fragile states. As shown in Table A3.1, including in the World Bank's list of fragile states shows a relationship with lower statistical significance. More curiously, using the Failed States Index shows no relationship. Keeping in mind that earlier findings also showed a relationship, our interpretation is that political stability appears to be fairly robust, but it should be kept in mind that results vary depending on which countries are considered more or less stable, and this involves judgment in particular for countries at risk of instability rather than manifestly instable or experiencing conflict. Also, as noted in World Bank 2012, some post-conflict countries though still rated as 'fragile' have been able to make significant progress on PFM reforms; and the range of 'fragile' or low stability states includes countries with varying other characteristics, such as income levels.

A shift in the regime type rating from the lowest score of 0 to the highest score of 10 (i.e. from fully authoritarian to fully democratic) would increase the average PEFA score by only 0.32 points on the 1-4 PEFA scale.[41] A growing literature documents that a variety of development outcomes do not correlate with the regime type. The assumed reason for this is that there are high levels of variation among (more) democratic as well as among (more) authoritarian regimes – with some cases in each group performing well and others badly.

As noted in section 4.2, we explore in addition whether the regime has an impact particularly on the external scrutiny and audit (dimension 6 of PEFA), which could be stronger than that for PFM systems overall. In particular, the external audit function and follow-up by the legislature may be more developed in countries with more democratic regimes, both because parliament can be expected to play a more independent role and because democratic governments can be presumed to enable an independent and effective external oversight body. However, we do not find a statistically significant relationship using either Freedom House or Polity IV data as a

---

columns 4 and 5). We also checked whether aid dependency is a factor for low income countries where aid dependency may be a more important driver of PFM reform. But the coefficient remained insignificant in regressions run separately for these countries.

[41] The results hold similar when using Polity IV instead of Freedom House (see Table A3.1, column 3) – both have small magnitudes of around 0.03.

measure for regime and PEFA dimension 6 scores as the dependent variable (see Annex 4, columns 7 and 8).

The effect of programmatic parties is positive as expected. Countries with a set of programmatic parties as compared to those with no programmatic parties will, on average, have a PEFA score that is 0.36 points higher. The relationship could be through greater coherence in governance that programmatic parties may bring, as well as due to the fact that programmatically based governments may have clearer policy priorities and will make greater efforts to pursue them, including pursuing better PFM systems. However, it is also important to note that the relationship we find is 'lose' and only significant at the 10 percent level, and thus is significantly weaker compared to the relationship between the presence of programmatic parties and the success of World Bank projects supporting public sector reforms that Keefer and Cruz (2010: 26) report, and more likely to be influenced by which countries are included, and how specific countries and parties are coded.

We graphically show these relationships using partial scatterplots in Figure 6 below.[42] The slope coefficients and standard errors are the same as the regressions reported in Table 3. The six graphs show the partial relationship between the quality of PFM systems (i) tax revenue to GDP, (ii) aid dependency, (iii) political stability, (iv), the nature of the political regime and (v) the presence of programmatic political parties, respectively, holding constant log GDP per capita, growth rates, population, SIDS, and resource dependency.

---

[42] These scatterplots are 'partial' regression plots because they show the partial correlation of PFM quality and the X variable of interest, after removing the linear effects of other independent variables in the model. Note that the X & Y axis cannot be interpreted as the values of PEFA scores or the X variable of interest because the values in the graphs are the residuals of these relationships. The vertical axis is the residuals from regressing PFM quality on the other X variables. The horizontal axis is the residuals from regressing the X variable of interest on all the other X variables.

	(1)	(2)	(3)	(4)	(5)	(6)
Table 3. Cross-Section Analysis: Average PEFA Scores and Country Characteristics						
GDP per capita (log)	0.2192***	0.1669***	0.2478***	0.1843***	0.1944***	0.1865***
	(0.0449)	(0.0582)	(0.0586)	(0.0449)	(0.0446)	(0.0566)
Growth (per capita)	0.0325***	0.0472***	0.0319***	0.0317***	0.0360***	0.0372***
	(0.0105)	(0.0144)	(0.0112)	(0.0097)	(0.0111)	(0.0103)
Population (log)	0.0556**	0.0502*	0.0648**	0.0941***	0.0558**	0.0495*
	(0.0276)	(0.0294)	(0.0293)	(0.0315)	(0.0271)	(0.0298)
SIDS	-0.2755*	-0.2621*	-0.2667*	-0.2647*	-0.3436**	-0.2858*
	(0.1428)	(0.1484)	(0.1440)	(0.1374)	(0.1419)	(0.1497)
Resource	-0.3816***	-0.2268**	-0.3684***	-0.3333***	-0.3233***	-0.3373***
	(0.0914)	(0.0949)	(0.0940)	(0.0883)	(0.0939)	(0.0955)
Tax		0.0137*				
		(0.0073)				
Aid			0.0025			
			(0.0042)			
Political Stability				0.0524***		
				(0.0196)		
Regime					0.0315*	
					(0.0173)	
Programmatic Parties						0.3573*
						(0.2044)
Observations	112	93	111	112	112	102
R-squared	0.40	0.41	0.41	0.42	0.42	0.43

Notes: Robust standard errors in parentheses; statistical significance is indicated as: *** $p<0.01$, ** $p<0.05$, *$p<0.1$

Figure 6: Partial Scatterplots

PFM Quality and Tax/GDP

PFM Quality and Aid

PFM Quality and Stability

PFM Quality and Regime
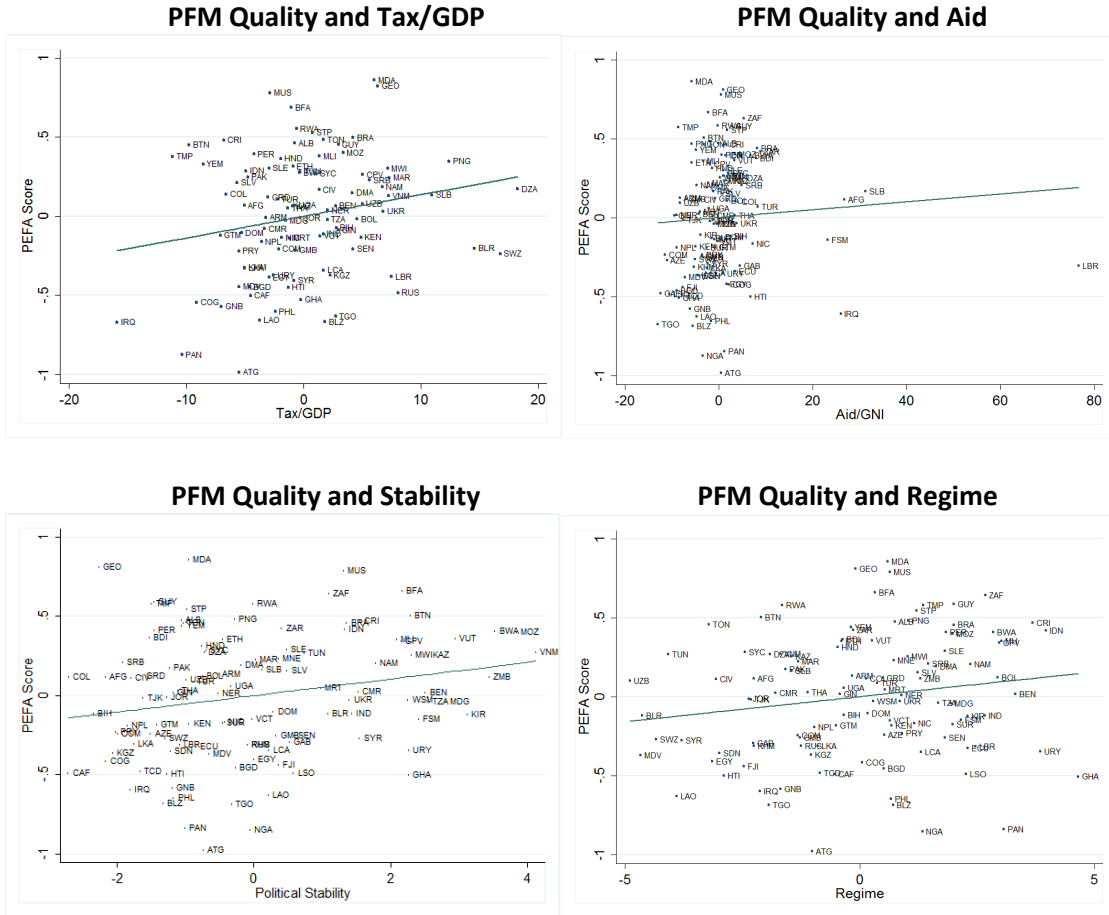
PFM Quality and Programmatic Parties
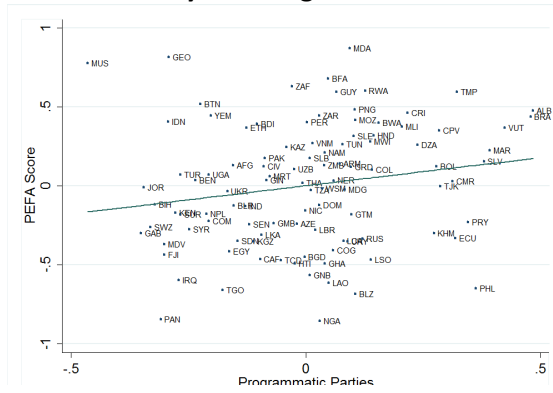
To further check robustness of the main findings, the regression in column (1) was run using CPIA-13 instead of average PEFA scores as a dependent variable (see column 12 of Table A3.1). Overall, the country variables in this regression account for only 32 percent of the variance in PFM quality as compared to 40 percent with PEFA. Similarly, we find that income and population have a positive relationship with PFM quality, and resource dependency a negative.

The biggest change relates to growth per capita and SIDS, which are no longer statistically significant in the regression using CPIA-13 as the dependent variable. These findings do not necessarily invalidate the results in the PEFA regression, but rather demonstrate the sensitivity of the model to the countries and specific years included. Whereas the regression with PEFA scores as the dependent variable (Table 3 column 1) includes 112 countries, for the CPIA-13 regression the number of countries increases to 126 (Table A3.1, column 12). For some countries there is also a shift in the years covered, as we use the 2012 CPIA rating for all, while for PEFA we use the most recent assessment available; this may affect the findings on the relationship with growth in particular.

Overall, these findings provide some boundaries around the expectations about which countries are likely to have or to develop in the short to medium term higher quality PFM systems. They can also help us to assess whether we consider a country to be under-performing or over-performing relative to other low or middle income countries included in the sample. At the same time, more than half of the variation in PFM quality remains unexplored and requires further investigation. This suggests that reform prospects are not closely bound by country characteristics – an important finding relative to our overall interest in whether particular reform approaches and political economy factors are likely to matter. We return to the overall implications in the conclusion. As a next step we pursue further how changes in country characteristics are associated with changes in PFM performance through looking at changes within countries over time.

5.2 Macro-Level Country Characteristics and PFM Quality: First Differences

While we found several macro-level country characteristics to be correlated with PFM quality in the first model, the single cross-section of data offered only inter-country (across) variation. One problem with this relationship is that there may be omitted variables that matter for both PFM performance and the country characteristics variables, which due to their exclusion are biasing the results. For example, the observed relationship between the economic growth and PFM quality at the cross-country level could be due to some unobserved variable which influences both. The exploration of within-country changes over time using the first differences method corrects for (time invariant) omitted variables. Furthermore, this line of investigation gives an indication of whether country characteristics should be taken into account when forming expectations about PFM quality improvement, for example at the time of the design of PFM reforms.

Since the length of the time interval between PEFA assessments varies by country, we cannot run a fixed-effects regression. Rather we manually compute the per year change in PEFA scores and the per year change over the same period in the variables capturing country

characteristics.[43] This allows us to relate changes in PFM quality to changes in country characteristics. Specifically we are asking whether, within a country, if its characteristics change, then how much is PFM quality expected to change? For the three characteristics that are typically not subject to year-to-year changes (being a SIDS, resource dependent, or fragile) we use a dummy, and hence measure whether countries with such characteristics show any significant difference in the rate of change of their PFM systems.

As a starting point, we take a look at the changes in PEFA scores across countries. Figure 7 presents a distribution of PEFA scores with the initial score on the X-axis and the most recent score on the Y-axis. The scatter plot shows a positive correlation across the 49 countries for which at least two PEFA reports are available -- with those above the 45-degree line showing improvements in the quality of their PFM systems and those countries below the 45-degree line showing declines. However, the slope of the linear fit is smaller than 45 degrees, suggesting that countries with a lower PEFA score at the outset end up with larger improvements in PFM quality than countries that start out with a higher score.

Figure 7. PFM Quality
Earliest vs. Latest Aggregate PEFA Scores
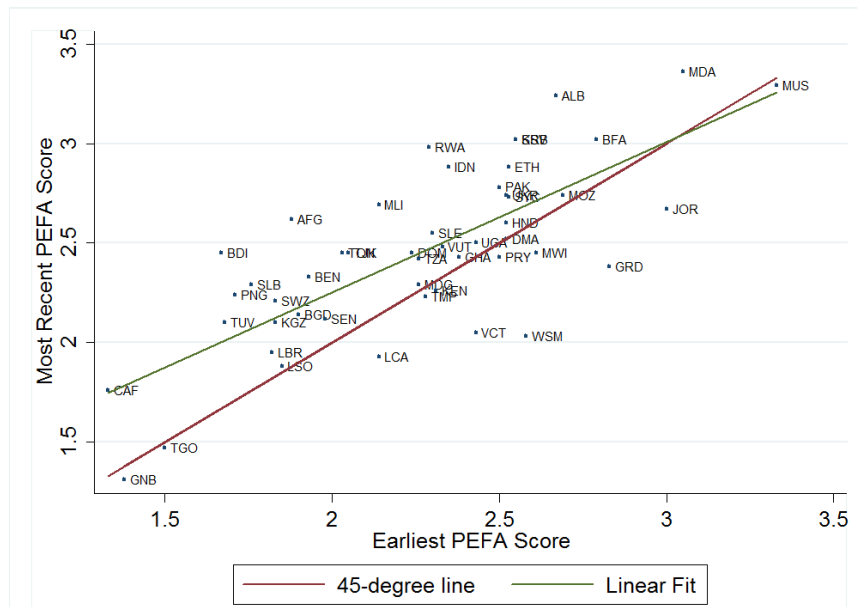
In order to see whether changes in PEFA scores over time are more pronounced in specific areas of PFM systems, we also examine the change in PEFA scores across the six sub-dimensions. Figure 8 displays two box plots for the 'first' and 'most recent' PEFA scores of our sample countries for which two or more assessments are available. Except for policy-based budgeting (dimension 3), the PEFA dimension scores have improved over time, with comprehensiveness and transparency (dimension 2) improving most. But the pace of improvement has not been

---

[43] Observations of country characteristics for the initial and subsequent PEFA assessments were not available in all cases. In such cases, data were intrapolated using observations for the country characteristics within 3 years of the PEFA assessment date.

uniform across countries. For a number of dimensions, the range of the middle 50 percent (indicated by the box) has widened, in particular for budget credibility (dimension 1) and external scrutiny and audit (dimension 6).

Figure 8. Aggregate PEFA Scores by Sub-Dimension
First and Last PEFA
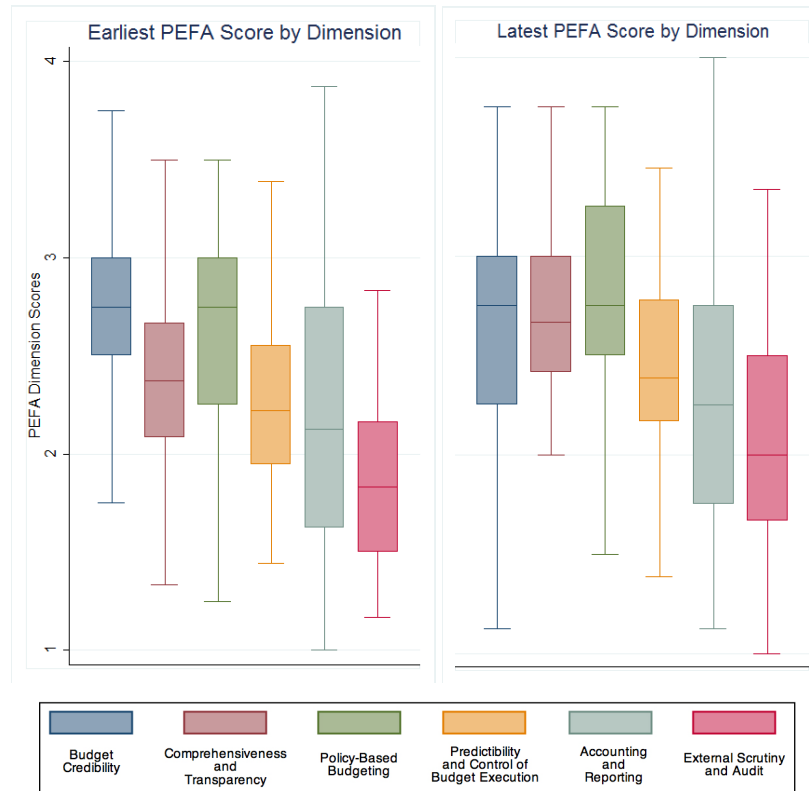
Similar to the analysis of the preceding section, we next examine the relationship between income and PFM quality using the first differences method. We estimate our results by running the average per year change in PEFA scores against the average annual rate of growth in GDP per capita across countries. Figure 9 below depicts this bi-variate relationship, which shows that higher growth countries, on average, show a greater increase in PFM compared to poorer countries over time (estimate = 1.2, standard error = 0.7).[44] However, the relationship is statistically significant only at the 10 percent confidence level, and the slope of the regression line as well as the large variance around the fitted line (similar to Figure 3 above) indicates that the impact of higher growth on improvements in PFM quality is very small.

We can again interpret these results by comparing two countries. By examining the Dominican Republic and Serbia – which have similar per capita income levels -- with 4.0 and 0.8 percent average annual growth, respectively, this model tells us that the difference between a fast growing and slow growing country of about 3 percentage points implies an average annual

---

[44] Timor-Leste is excluded from the Graph, as it is an outlier on the change in income level.

improvement in PEFA scores of 0.036. Over a three-year period, this would lead to an increase of 0.108 in the PEFA score.[45] While these results highlight that the levels equation predicts a much smaller change in PFM quality than the differenced equation, both models tell us that income is positive and significant.[46]

Figure 9. Change in Income Level Change and in PFM Quality
Between Earliest and Most Recent PEFA Scores
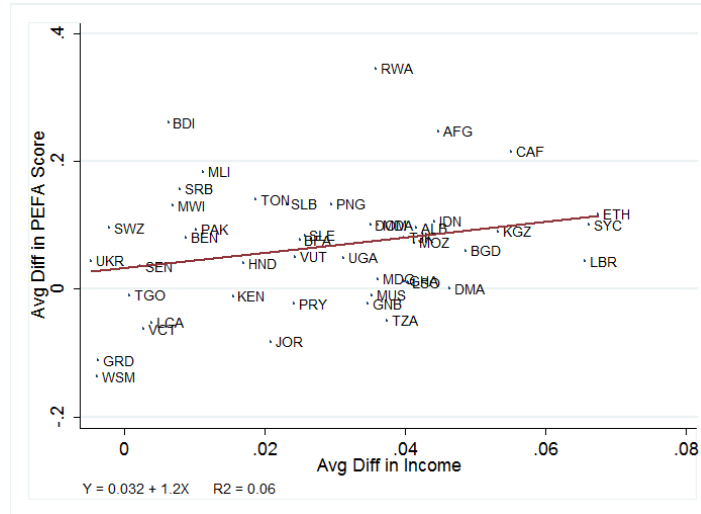
We confirm our cross-sectional findings when we include additional country characteristics variables into our first-differences model, which is presented in Table 4.[47] Except for SIDS, country characteristics and changes in country characteristics, including economic growth and population size, are weakly related to improvements in PFM quality. Instead, the quality of PFM systems at the beginning of the period is the key determinant of the change that can be expected in subsequent period (see column (2)).

Consistent with our findings in Figure 7, the coefficient for initial PFM quality has a negative sign, suggesting that the scope for improvements in PFM quality is larger in countries with relatively weak PFM systems as compared to higher-performing countries over time. The edge provided by weaker PFM quality is substantial – with a coefficient estimate of about 0.08, which indicates that the difference in PEFA scores between two countries, one with an average score of B (2 in the numerical conversion) and another with the lowest score of D (equivalent to 4), would see the difference between their scores decline from 2 to 1.84 within a year, a decline of 8 percent. In itself, this is plausible—countries with higher initial PFM quality may have less need

---

[45] The Dominican Republic and Serbia have an annual average change in GDP per capita of 3.97 and 0.78 percent, respectively (each over a three-year period between first and latest PEFA assessments).

[46] Note that we cannot do a precise comparision because there is a large difference in the sample of countries, with a much smaller number of countries included in model 2. In addition, the difference in findings may be driven by a sample bias towards countries that are more inclined to embark on PFM reforms as these countries have chosen to undertake and publish a second PEFA assessment.

[47] Paucity of data counsel cautions in the interpretation of these results. As more repeat PEFAs are undertaken, this preliminary analysis of changes in PFM system quality needs to be updated and the association with country characteristics investigated further.

34

for further improvements, and may find it more difficult to achieve them. But this also suggests that the scope for PFM reform is only to a limited extent constrained by country characteristics, which are largely outside the (short-term) control of decision-makers in government and other PFM reform stakeholders. This is also consistent with the starting point for the follow-up work under the study, which is that country-specific constellations and dynamics critically influence when and to what extent a strengthening of PFM systems happens.

To check the robustness of the main findings, we run the first-differences model using CPIA-13 instead of average PEFA scores as a dependent variable (see Table A3.2).[48] Due to the higher number of countries available (124 countries as compared to 40), our regression fit improves significantly with country characteristics and changes in country characteristics accounting for around 40 percent of the variance in PFM quality, which is similar to that in the cross-sectional analysis. Having lower initial PFM quality has a robustly significant effect, similar to the results when using PEFA data. The biggest change is that using CPIA data a higher rate of population change emerges as having a significant negative effect on CPIA-13 as the dependent variable, suggesting that countries with high population growth have been among those achieving more limited PFM reforms. Being a SIDS and being resource dependent assume a small negative effect, significant at the 10 percent level using this wider set of countries, which is consistent with the findings from the cross-sectional analysis. Changes in GDP per capita appear to have a larger effect than the cross-sectional results with regards to growth. Among the political variables, the CPIA results show a small negative effect of being fragile and a positive effect of having programmatic parties.

---

[48] The analysis is based on a comparison between CPIA data for 2005 and 2012, i.e. broadly the same time-frame as that for which PEFA data is available; but with a wider country coverage.

Table 4. First-Differences Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP per capita (percent change)	0.6931	0.7424	0.5837	0.7898	1.0917*	0.7046	0.2296	0.7640
	(0.5939)	(0.5695)	(0.6579)	(0.6513)	(0.6344)	(0.5760)	(0.6083)	(0.6211)
Population (percent change)	0.8019	-0.4527	-0.3918	-0.4633	-1.1561	-0.5722	-0.4310	-0.4900
	(1.1686)	(1.2903)	(1.7871)	(1.6397)	(1.9450)	(1.3005)	(1.2796)	(1.4811)
Resource (dummy)	0.0132	0.0095	0.0170	0.0006	0.0054	0.0085	0.0423	0.0100
	(0.0493)	(0.0404)	(0.0418)	(0.0735)	(0.0854)	(0.0390)	(0.0378)	(0.0426)
SIDS (dummy)	-0.0614**	-0.0665**	-0.0481	-0.0691**	-0.0810*	-0.0651**	-0.0682**	-0.0664**
	(0.0268)	(0.0262)	(0.0390)	(0.0323)	(0.0415)	(0.0274)	(0.0322)	(0.0268)
Initial PFM quality (PEFA) (level)		-0.0796**	-0.0663*	-0.0691*	-0.0796**	-0.0823**	-0.0938***	-0.0799**
		(0.0322)	(0.0360)	(0.0385)	(0.0339)	(0.0325)	(0.0339)	(0.0332)
Initial GDP per capita (level in log)			0.0079					
			(0.0265)					
Initial regime type			-0.0124					
			(0.0088)					
Tax (percentage point change)				-0.0194				
				(0.0209)				
Aid (ODA) (percentage point change)					-0.0040			
					(0.0127)			
Regime (Freedom House, percent change)						-0.0282		
						(0.0693)		
Programmatic Parties (percent change)							-0.0875	
							(0.0667)	
Political Stability (percent change)								-0.0040
								(0.0571)
Observations	47	47	47	36	33	47	41	47
R-squared	0.15	0.25	0.30	0.21	0.31	0.26	0.33	0.25

5.3 The Effects of PFM Quality: Exploring Fiscal Outcomes and Public Service Delivery

As set out in the preceding sections, apart from inquiring into the drivers of PFM performance, an important issue is whether and how PFM performance is associated with outcomes. As discussed, we hypothesize that expected outcomes from better PFM quality are: improved aggregate fiscal discipline, improved allocative efficiency, and improved operational efficiency. Attempts at generating cross-country empirical evidence are still rare; and the challenges with regards to identifying relevant proxies and finding available cross-country data are still considerable. Therefore, the evidence presented here should be recognized as preliminary, and as an attempt to begin filling a gap.

Table 5 below shows the regression output between the quality of PFM systems and the three key objectives: overall fiscal balance and the aggregate budget execution rate (as proxies for aggregate fiscal discipline), composition of expenditure out-turn compared to original approved budget (as a proxy for allocative efficiency), and cost-effectiveness of health expenditure and of education expenditure (as proxies for operational efficiency). OLS is used to explore the relationships, controlling for log GDP per capita. The partial scatterplots are shown in Figure 10 below.

We find that the relationship between the quality of PFM and aggregate fiscal discipline when measured by the level of government primary net lending/borrowing is not statistically significant, but we do find a positive and significant relationship when we examine the association with the overall budget execution rate. We find a coefficient of 0.683 (significant at the 99 percent level and with a standard error of 0.2) for the countries included in the analysis. A plausible link between the two is that countries with better PFM systems are more likely to stick to their annual planned budgets. It is possible that the lack of relationship with deficit levels is related to the time period, that is, the fact that many PEFA assessments were done as part of the process toward debt relief and during the global financial crisis, which has prompted larger deficits in many countries, including those with stronger PFM systems. The limited number of observations (56) also makes it more difficult to establish statistical relationships.

With regards to PI-2 measuring budget credibility in terms of sector allocations being aligned with original allocations, we find a coefficient of 0.629 and standard error of 0.22 (using simple OLS, controlling for GDP per capita and significant at the 99 percent level). This suggests that strengthening PFM systems indeed results in governments generating more realistic and credible budgets and executing them as planned in terms of how funds are allocated across main sectors.

We do not find evidence that health or education results relative to public sector spending are better in countries with stronger PFM systems, when controlling for GDP per capita. This is consistent with recent work on the effects of MTEFs, which finds that only the most developed form of an MTEF – a medium term performance framework (MTPF) – shows any significant correlation with operational efficiency as measured by the cost-effectiveness of public health expenditures (World Bank 2013: 48-50). How we can better measure impacts of PFM quality on service delivery capabilities remains a field for further investigation. This is of considerable interest, given the expectation that developing better PFM systems will yield pay-offs in terms

of improved service delivery; but some forms of intermediate data points may be needed that capture steps in the causal chain more directly influenced by PFM systems than final outcomes.[49]

Table 5: PFM Quality and Outcome Variables

Dependent Variable:	(1) Deficit	(2) PI-1	(3) PI-2	(4) Health	(5) Education[50]
PFM Quality	-1.8614	0.6830***	0.6408***	-0.6796	0.3896
	(1.5665)	(0.2160)	(0.2197)	(3.7362)	(4.2376)
GDP per capita (log)	0.9401	-0.0991	0.1766	9.0169***	-14.0817***
	(0.7393)	(0.1177)	(0.1318)	(1.5627)	(2.1150)
Observations	56	102	97	60	57
R-squared	0.05	0.10	0.18	0.45	0.52

Notes: Robust standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

One possible explanation for this lack of a significant relationship could be the fact that there is considerable variation between the strength of PFM systems (as reflected in PEFA indicators) and overall government effectiveness, as shown in section 4.1. If PFM systems are relatively strong, but government effectiveness remains limited otherwise, then potential beneficial effects on service delivery would 'dissipate'. However, at least for the sample for which data on expenditure effectiveness can be constructed, overall government effectiveness also shows no significant relationship with efficiency in service delivery.

It is important to keep in mind that the measures currently available are rather broad proxies. For example, actual spending levels reaching local levels and specifically, front-line service delivery agencies, and the quality of PFM at those levels may have more direct effects on spending efficiency, but no specific cross-country measures of these are currently available.[51]

---

[49] There is also the challenge that health or education outcomes depend not only on public funding but also on how much citizens spent privately out of pocket. User satisfaction with public facilities (relative to public sector spending) may be useful as an intermediate indicator, but available data is typically country specific rather than comparable across cases; while there may be opportunities for comparisons across time.

[50] We confirm our findings when we formulate the education efficiency variable using adult literacy rates as compared to primary school completion rates.

[51] In addition to using the aggregate PEFA scores constructed as described in section 4.1, we also checked for the relationship between any of the six PEFA subdimensions and the outcome variables; but these follow similar patterns as the aggregate scores or show no clear relationship.

Figure 10: Partial Scatterplots

PFM Quality and Fiscal Balance
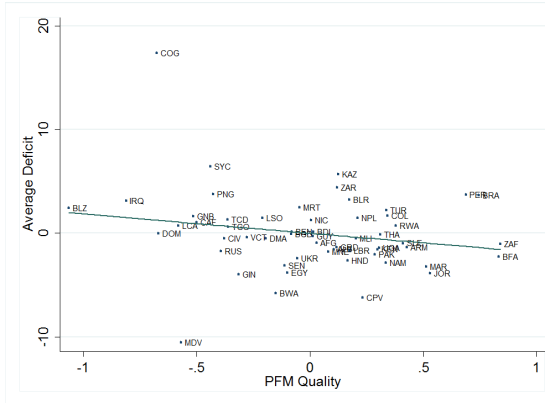
PFM Quality and PI-1
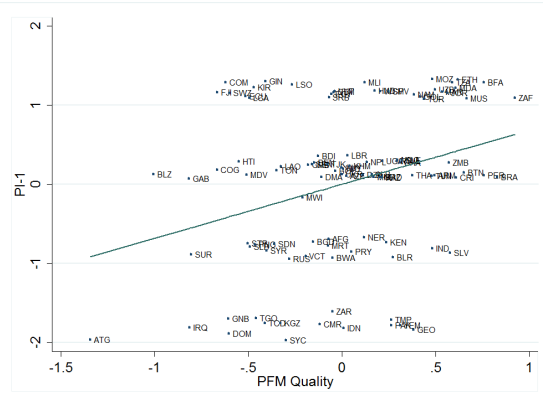
PFM Quality and P-2
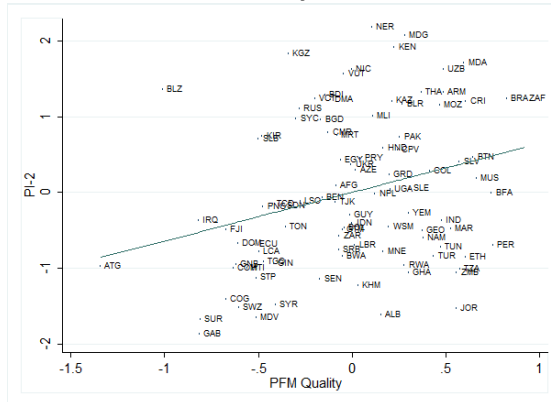
PFM Quality and Education Cost-effectiveness
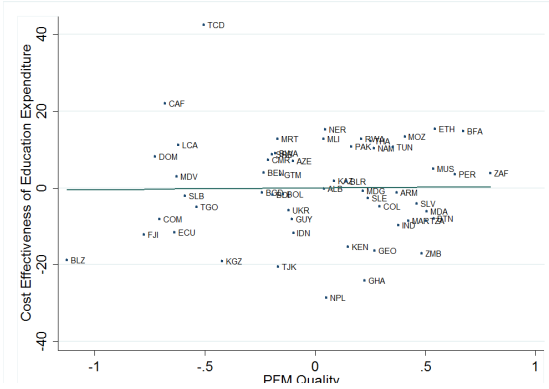
PFM Quality and Health Cost-effectiveness
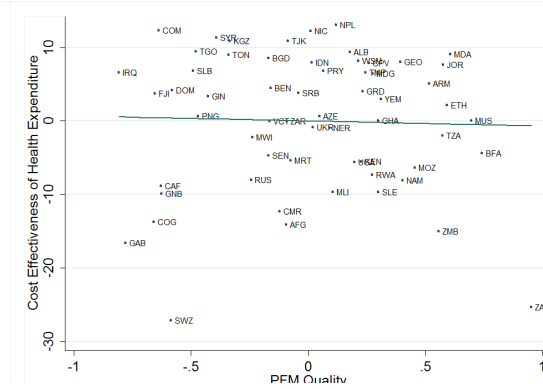
6. Conclusion

This paper explored two relationships, one between country characteristics and PFM quality and second, that between the quality of PFM systems and expected outcomes. As introduced at the outset, it is part of a wider effort at integrating considerations about non-technical drivers more explicitly into PFM reform approaches and doing so in ways that can inform efforts to strengthen PFM systems. The intention of this work has been two-fold: firstly, to revisit and build on existing evidence about cross-country patterns, and secondly, to add to early explorations of whether improvements in PFM quality indeed have expected effect on outcomes.

As set out in section 1, the guiding question is whether and to what degree country characteristics influence PFM quality and prospects for PFM strengthening. The paper investigated the following set of factors: economic factors (level, growth and resource dependency), population (including SIDS), levels and sources of revenue (tax revenue to GDP, the level of aid relative to GNI), and three macro-political characteristics – political stability, regime type, and the presence of programmatic parties. The more closely such country characteristics are associated with differences in PFM systems performance, the less need there is to think about 'smart' reform approaches; but also the less opportunity there would be for countries to pursue PFM reforms if they have unfavorable characteristics. Methodologically, a key limitation is the issue of causality, which cannot be fully addressed with the types of data and information available. In particular, endogeneity and omitted variables need to be kept in mind as key issues, and our results should be interpreted with due caution in this regard.

Overall, we find that the macro-level country characteristics we explore jointly explain about 40 percent of the variation in PFM quality across countries. The strongest associations that we find are with income level (positive), and with being a small island state, and resource dependency (negative). We also find an association with political stability and with having programmatic parties (both positive), albeit for the latter weaker in terms of statistical significance; and with general population and growth (positive but with a very limited impact). The level of tax revenue and regime type also show an association, but these results are borderline and in particular for revenue not robust. We also observe a strong negative association between HIPC countries and the quality of PFM, while in this regard, we assume that the more likely direction of causation is from weak PFM systems to becoming highly indebted.

Therefore, country characteristics do not appear to have a strong predetermining effect for most countries, and particular reform approaches including political economy considerations are likely to matter for reform success. However, for countries that combine being low income with being small island states or with being dependent on natural resources (or both) these characteristics appear to be relatively stronger constraints (again, keeping in mind that our analysis does not prove a causal relationship).

This is consistent with our overall starting assumption that constellations of stakeholders, institutional, and structural factors and interaction dynamics – which to date are not well captured by quantitative indicators – are likely to significantly affect PFM reform efforts and outcomes, and that specific reform approaches matter. For example, constellations between these factors would determine what PFM reform efforts are being made in reaction to the experience of a fiscal crisis, rather than the crisis event as such being a strong predictor of

subsequent reform results. The variable included here that comes closest to capturing one aspect of these issues is the presence of programmatic parties, and our findings suggest that countries which combine political stability and structured parties tend to achieve better PFM performance.

Furthermore, the first-difference analysis suggests that countries with initially lower PFM quality show a higher rate of improvements over time, and this results holds when using PEFA as well as CPIA data. Being a small island developing state shows a small negative association, also being a resource rich country, albeit only when using CPIA data (with more observations). Faster GDP growth may have a positive, and faster population growth appears to have a negative association with improvements in PFM quality.

As better and more data become available, these findings would need to be revisited and validated. Overall, the results, if validated, are encouraging from a PFM reform perspective: reform prospects of relatively poor performers appear to be relatively promising, and the constraints from country characteristics are limited. Thus, the findings imply that an inquiry into specific reform dynamics is worthwhile, and that identifying strategically smart reform approaches may make a difference in the timing and the extent of PFM reforms achievable.

With regards to the effects of PFM performance, we find evidence that stronger performance results in better budget credibility in terms of overall budget execution rates as well as allocations across main functions/sectors, albeit not in lower deficits. This suggests a positive effect on allocative efficiency (assuming that budget plans reliably reflect policy priorities), and at least a partial effect on aggregate discipline, in the sense that governments are able to keep deficits within planned amounts. There is no clear evidence with regards to effects on operational efficiency (while noting that this is also the most challenging effect to measure). Government leaders mandating and pursuing PFM reforms are likely to do so with a view to achieving certain outcomes.[52] Hence, the observation of effects implies that leaders may feel that there is some return to the investments made in PFM improvements. However, at the same time, the absence of clearer effects on service delivery results poses a challenge. The observed 'disconnect' could be due to a variety of reasons – such as potential missing complementary state capacities, or measuring problems, or others – which need to be explored further.

In a second step of the wider effort, we intend to explore actual processes and prospects of PFM reforms for selected individual countries. These case examples will be informed by the assumptions resulting from this cross-sectional work, and will explore further how presumed drivers for PFM reforms matter within specific situations. Given the findings, we presume that a number of more case-specific factors and constellations may play an important role – i.e. even countries with some similar characteristics may diverge substantially in terms of PFM reform results achieved due to these more specific factors. As part of further efforts under this workstream, case studies will explore the specific motivations of pursuing PFM reforms in the countries concerned, as well as how stakeholders in those cases view the results and impacts achieved to date; and how this influences the prospects and motivation for further efforts.

---

[52] Broadly, we assume that political leaders 'mandate' or endorse PFM reforms and the intended outcomes, while leaders within the executive would actually pursue the implementation of reforms/PFM strengthening.

Finally, as pointed out with regards to various aspects of this analysis, further testing and exploration of many of the issues will be valuable; and there are also a number of aspects, in particular related to the effects of PFM on operational efficiency, for which more specific data comparable across countries would be needed, and will hopefully become increasingly available in the coming years.

ANNEX 1: Summary Statistics

Table A1.1 Summary Statistics (PFM and Country Characteristics)

Variable	Obs	Mean	Std. Dev.	Min	Max
PEFA	120	2.4	0.51	1.31	3.55
GDP per capita (log)	112	8.03	0.93	5.82	9.93
Growth (per capita)	118	3.3	3.02	-1.34	20.28
Population (log)	120	15.53	2.14	9.19	20.87
SIDS	120	0.24	0.43	0	1
Resource	120	0.21	0.41	0	1
Tax	94	16.81	6.75	0.65	37.8
Aid	116	9.16	13.32	-0.11	106.58
Political Stability (WGI)	119	3.75	2.50	0.03	9.82
Regime (Freedom House)	118	5.47	2.72	0	10
Programmatic Parties	105	0.27	0.24	0	1
Additional or alternative variables					
CPIA-13	135	3.41	0.72	1	5.5
Fragility	120	0.23	0.42	0	1
FS Index (Foreign Policy)	83	83.09	14.44	41.2	111.1
Regime (Polity IV)	102	6.53	2.69	0.5	10
Aid (Technical Cooperation)	81	1.16	1.14	0.03	4.74
Aid (CPA)	81	9.86	24.93	0.05	210.18
HIPC	120	0.05	0.22	0	1
Fiscal Shock	49	0.76	1.13	0	4
Growth Shock (4 year lag)	120	0.1	0.3	0	1
Growth Shock (5 year lag)	120	0.03	0.16	0	1

Note: These summary statistics are based on the year of the country's most recent PEFA assessment, except for CPIA-13, which was based on 2011.

Table A1.2. Summary Statistics (PFM and Outcomes)

Variable	Obs	Mean	Std. Dev.	Min	Max
PEFA	106	2.30	0.52	1.33	3.55
GDP per capita (log)	106	8.04	0.97	5.77	9.91
Deficit	56	-1.10	3.81	-11.41	16.31
PI-1	102	2.81	1.01	1	4
PI-2	97	2.38	1.06	1	4
Health efficiency	60	79.90	12.44	56.71	98.18
Education efficiency	57	44.56	18.36	14.23	95.04

Note: Summary statistics were formulated based on the year in which a country's first PEFA assessment was performed

ANNEX 2: Correlations between Independent Variables

Table A2.1 Correlation Statistics

	Income	Growth	Population	SIDS	Resource	Tax	Aid	Regime	Political Stability	Programmatic Parties
Income	1									
Growth	0.1413	1								
Population	-0.1514	0.0695	1							
SIDS	0.0703	-0.0946	-0.5998	1						
Resource	0.0383	-0.1637	0.0436	0.0296	1					
Tax	0.4299	0.0326	-0.1132	0.0281	0.0078	1				
Aid	-0.5291	0.1151	-0.1539	0.0887	-0.0563	-0.1546	1			
Regime	0.2743	-0.0691	-0.24	0.3342	-0.2246	0.0797	-0.0594	1		
Political Stability	0.3787	0.1681	-0.5348	0.2381	-0.1155	0.3326	-0.0881	0.3957	1	
Programmatic Parties	0.4362	-0.0491	0.0488	0.0266	-0.0915	0.223	-0.2789	0.3697	0.1895	1

ANNEX 3: Robustness Checks and Additional Quantitative Analysis

We use alternative sources and methodologies of our dependent and independent variables in an attempt to minimize measurement error.[53] The alternative sources as follows:

Variable	Main source	Alternatives
PFM Quality	PEFA	CPIA-13
Income	GDP per capita, PPP (constant 2005 international $)	GDP per capita (constant 2005 US$)
Tax	Tax/GDP (IMF FAD's database)	Total Revenue/GDP (IMF FAD's database)[54]
Aid	ODA	Country Programmable Assistance and Technical Cooperation
Political stability/fragility	World Bank's World Governance Indicators (Political Stability)	World Bank's Fragile States list (fragility) and Foreign Policy's Failed States index (FS index)
Regime	Freedom House	Polity IV
Education	Primary completion rate	Adult literacy rate

The following additional country characteristics are also analyzed:

Additional variables			
Fiscal Shocks	Fiscal shocks may trigger efforts to strengthen PFM systems so as to reduce the probability of future shocks	Caceres and Kochanova, IMF 2012	Discrete variable for number of fiscal shocks experienced (in a five-year period)
Growth Shock	Similarly, growth shocks may trigger efforts to strengthen PFM systems so as to improve resilience and the fiscal response to future shocks	World Bank WDI	Dummy variable for all countries that have a 'growth shock' (i.e. when the annual percentage growth rate of GDP is less than negative 2%)
HIPC	Debt relief was provided on the basis of improvements in PFM systems quality in the run up and after countries reached the decision point and donors provided capacity building support for achieving agreed goals. HIPCs may therefore show larger improvements in PFM systems quality than other countries.	World Bank	Dummy variable for countries who have passed the decision point under the HIPC and DRMI initiatives

---

[53] Further checks not included in the report but which confirm our findings include different lengths of lags (and no lags) and alternative aggregation techniques for PEFA as well as the inclusion and exclusion of LICs and MICs.

[54] If data is available, includes grants and social contributions. See the Methodology in the IMF FAD database for more information.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable:	PEFA	PEFA	PEFA	PEFA	PEFA	PEFA	PEFA	PEFA	PEFA	PEFA	PEFA	CPIA
GDP per capita (PPP, log)		0.1923***	0.2331***	0.3448***	0.2584***	0.1801***	0.1952***	0.3193***	0.2048***	0.2184***	0.1910***	0.3515***
		(0.0502)	(0.0444)	(0.0727)	(0.0539)	(0.0485)	(0.0675)	(0.0657)	(0.0461)	(0.0450)	(0.0450)	(0.0573)
Growth (per capita)	0.0341***	0.0451***	0.0362***	0.0286***	0.0307***	0.0307***	0.0323**	0.0272	0.0352***	0.0299***	0.0306***	0.0068
	(0.0109)	(0.0147)	(0.0107)	(0.0099)	(0.0106)	(0.0104)	(0.0123)	(0.0205)	(0.0109)	(0.0100)	(0.0100)	(0.0229)
Population (log)	0.0536**	0.0495*	0.0441	0.0965***	0.0688**	0.0462*	0.0538	0.0390	0.0664**	0.0540*	0.0452	0.0802**
	(0.0268)	(0.0286)	(0.0291)	(0.0335)	(0.0329)	(0.0277)	(0.0343)	(0.0347)	(0.0290)	(0.0277)	(0.0279)	(0.0328)
SIDS	-0.3649**	-0.2986**	-0.2833*	(0.4117)	(0.4166)	-0.2687*	-0.3270*	-0.4238**	(0.2369)	-0.2728*	-0.3122**	(0.0164)
	(0.1487)	(0.1411)	(0.1584)	(0.2948)	(0.2869)	(0.1370)	(0.1811)	(0.1986)	(0.1456)	(0.1435)	(0.1414)	(0.1513)
Resource	-0.3786***	-0.3168***	-0.3539***	-0.3957***	-0.4118***	-0.3602***	-0.4332***	-0.4076**	-0.3970***	-0.3676***	-0.3808***	-0.4720***
	(0.0945)	(0.0955)	(0.0950)	(0.1018)	(0.1058)	(0.0909)	(0.1087)	(0.1746)	(0.0889)	(0.0893)	(0.0892)	(0.1307)
GDP per capita (cons, log)	0.1915***											
	(0.0402)											
Revenue		0.0075										
		(0.0049)										
Regime (Polity IV)			0.0283*									
			(0.0155)									
Aid (Technical Cooperation)				0.0887								
				(0.0650)								
Aid (CPA)					(0.0003)							
					(0.0011)							
Fragility (dummy)						-0.1778*						
						(0.1000)						
FS Index (Foreign Policy)							(0.0047)					
							(0.0050)					
Fiscal Shocks								(0.0303)				
								(0.0454)				
Growth Shock (4 year lag)									0.2297*			
									(0.1371)			
Growth Shock (5 year lag)										-0.3962**		
										(0.1798)		
HIPC											-0.4001***	
											(0.1262)	
Observations	115	96	98	80	80	112	80	49	112	112	112	126
R-squared	0.398	0.409	0.443	0.423	0.407	0.434	0.416	0.558	0.415	0.411	0.429	0.321

Notes: Robust standard errors in parentheses; statistical significance is indicated as: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	CPIA	CPIA	CPIA	CPIA	CPIA	CPIA	CPIA	CPIA
GDP per capita (percent change)	0.3427	0.3144*	0.4248**	0.3527*	0.3627*	0.3120*	0.3145*	0.1347
	(0.2330)	(0.1756)	(0.1813)	(0.1892)	(0.1846)	(0.1765)	(0.1828)	(0.1896)
Population (percent change)	-1.0072*	-1.8032***	-1.3101**	-1.7163***	-1.8547***	-1.8047***	-1.8397***	-1.9844***
	(0.5370)	(0.4321)	(0.5985)	(0.4360)	(0.4802)	(0.4354)	(0.4400)	(0.4249)
Resource (dummy)	(0.0101)	-0.0200*	-0.0211*	(0.0125)	-0.0235**	-0.0191*	-0.0179*	-0.0165
	(0.0124)	(0.0105)	(0.0108)	(0.0128)	(0.0106)	(0.0106)	(0.0103)	(0.0104)
SIDS (dummy)	(0.0125)	-0.0203*	-0.0283**	(0.0184)	-0.0223*	-0.0203*	(0.0078)	-0.0158
	(0.0126)	(0.0110)	(0.0115)	(0.0120)	(0.0114)	(0.0111)	(0.0119)	(0.0111)
Initial PFM quality (CPIA-13) (level)		-0.0498***	-0.0562***	-0.0487***	-0.0480***	-0.0499***	-0.0477***	-0.0472***
		(0.0068)	(0.0069)	(0.0076)	(0.0074)	(0.0068)	(0.0068)	(0.0069)
Initial GDP per capita (level in log)			0.0050					
			(0.0072)					
Initial regime type			0.0034*					
			(0.0020)					
Tax (percentage point change)				0.0149				
				(0.0093)				
Aid (ODA) (percentage point change)					0.0008			
					(0.0041)			
Regime (Freedom House, percent change)						0.0179		
						(0.0351)		
Programmatic Parties (percent change)							0.0271**	
							(0.0133)	
Political Stability (percent change)								0.0555*
								(0.0285)
Observations	124	124	124	89	116	124	116	120
R-squared	0.086	0.390	0.411	0.444	0.373	0.391	0.406	0.415

Notes: Robust standard errors in parentheses; statistical significance is indicated as: *** p<0.01, ** p<0.05, *p<0.1

ANNEX 4: PEFA Dimensions Analysis

Table A4.1 Cross-Sectional Results for Country Characteristics and PEFA sub-dimensions

dependent variable:	(1) Dim 1	(2) Dim 2	(3) Dim 3	(4) Dim 4	(5) Dim 5	(6) Dim 6	(7) Dim 6	(8) Dim 6
GDP per capita (log)	0.3855***	0.2409***	0.0819	0.2101***	0.3165***	0.1592**	0.1944***	0.2331***
	(0.0666)	(0.0589)	(0.0496)	(0.0479)	(0.0768)	(0.0620)	(0.0446)	(0.0444)
Growth (per capita)	0.0256	0.0428***	0.0216	0.0229**	0.0341**	0.0277	0.0360***	0.0362***
	(0.0174)	(0.0132)	(0.0174)	(0.0108)	(0.0144)	(0.0181)	(0.0111)	(0.0107)
Population (log)	-0.0039	0.1081***	0.0224	0.0257	0.0923**	0.0047	0.0558**	0.0441
	(0.0496)	(0.0340)	(0.0497)	(0.0346)	(0.0414)	(0.0368)	(0.0271)	(0.0291)
SIDS	-0.5141**	-0.1587	-0.2940	-0.3944**	-0.1697	-0.5081***	-0.3436**	-0.2833*
	(0.2329)	(0.1919)	(0.2317)	(0.1561)	(0.2059)	(0.1791)	(0.1419)	(0.1584)
Resource	-0.3364**	-0.5796***	-0.3120**	-0.2327**	-0.4287***	-0.1141	-0.3233***	-0.3539***
	(0.1424)	(0.1363)	(0.1380)	(0.0889)	(0.1415)	(0.1431)	(0.0939)	(0.0950)
Regime (FH)							0.0315*	
							(0.0173)	
Regime (Polity IV)								0.0283*
								(0.0155)

ANNEX 5: List of countries[55]

Country	Model 1A	1B	Country	Model 1A	1B	Country	Model 1A	1B
Afghanistan	X	X	Guatemala	X		Paraguay	X	X
Albania	X	X	Guinea	X		Peru	X	
Algeria	X		Guinea-Bissau	X	X	Philippines	X	
Antigua and Barbuda	X		Guyana	X		Russian Federation	X	
Armenia	X		Haiti	X		Rwanda	X	X
Azerbaijan	X		Honduras	X	X	Samoa	X	X
Bangladesh	X	X	India	X		Sao Tome and Principe	X	
Belarus	X		Indonesia	X	X	Senegal	X	X
Belize	X		Iraq	X		Serbia	X	X
Benin	X	X	Jamaica	X		Seychelles	X	X
Bhutan	X		Jordan	X	X	Sierra Leone	X	X
Bolivia	X		Kazakhstan	X		Solomon Islands	X	X
Bosnia and Herzegovina	X		Kenya	X	X	South Africa	X	
Botswana	X		Kiribati	X		South Sudan	X	
Brazil	X		Kosovo	X	X	Sri Lanka	X	
Burkina Faso	X	X	Kyrgyz Republic	X	X	St. Lucia	X	X
Burundi	X	X	Lao PDR	X		St. Vincent and the Grenadines	X	X
Cambodia	X		Lesotho	X	X	Sudan	X	
Cameroon	X		Liberia	X	X	Suriname	X	
Cape Verde	X		Madagascar	X	X	Swaziland	X	X
Central African Republic	X	X	Malawi	X	X	Syrian Arab Republic	X	
Chad	X		Maldives	X		Tajikistan	X	X
Colombia	X		Mali	X	X	Tanzania	X	X
Comoros	X		Marshall Islands	X		Thailand	X	
Congo, Dem. Rep.	X		Mauritania	X		Timor-Leste	X	X
Congo, Rep.	X		Mauritius	X	X	Togo	X	X
Costa Rica	X		Micronesia, Fed. Sts.	X		Tonga	X	X
Cote d'Ivoire	X		Moldova	X	X	Tunisia	X	
Dominica	X	X	Montenegro	X		Turkey	X	
Dominican Republic	X	X	Morocco	X		Tuvalu	X	X
Ecuador	X		Mozambique	X	X	Uganda	X	X
Egypt, Arab Rep.	X		Myanmar	X		Ukraine	X	X
El Salvador	X		Namibia	X		Uruguay	X	
Ethiopia	X	X	Nepal	X		Uzbekistan	X	
Fiji	X		Nicaragua	X		Vanuatu	X	X
Gabon	X		Niger	X		Vietnam	X	
Gambia, The	X		Nigeria	X		West Bank and Gaza	X	
Georgia	X		Pakistan	X	X	Yemen, Rep.	X	
Ghana	X	X	Panama	X		Zambia	X	
Grenada	X	X	Papua New Guinea	X	X	Zimbabwe	X	

[55] This list covers the low and middle-income countries included in the regressiosn between PFM quality and GDP income per capita. Due to data limitations in some of the economic and political variables explored in Models 1A and 1B, several additional countries are dropped in subsequent regressions.

ANNEX 6: PEFA Dimensions and Indicators[56]

PFM-OUT-TURNS: Credibility of the budget

PI-1   *Aggregate expenditure out-turn compared to original approved budget*

PI-2   *Composition of expenditure out-turn compared to original approved budget*

PI-3   *Aggregate revenue out-turn compared to original approved budget*

PI-4   *Stock and monitoring of expenditure payment arrears*

KEY CROSS-CUTTING ISSUES: Comprehensiveness and Transparency

PI-5   Classification of the budget

PI-6   Comprehensiveness of information included in budget documentation

PI-7   Extent of unreported government operations

PI-8   Transparency of inter-governmental fiscal relations

PI-9   Oversight of aggregate fiscal risk from other public sector entities.

PI-10   Public access to key fiscal information

BUDGET CYCLE

(i) Policy-Based Budgeting

PI-11   Orderliness and participation in the annual budget process

PI-12   Multi-year perspective in fiscal planning, expenditure policy and budgeting

(ii) Predictability and Control in Budget Execution

PI-13   *Transparency of taxpayer obligations and liabilities*

PI-14   *Effectiveness of measures for taxpayer registration and tax assessment*

PI-15   *Effectiveness in collection of tax payments*

PI-16   Predictability in the availability of funds for commitment of expenditures

PI-17   Recording and management of cash balances, debt and guarantees

PI-18   Effectiveness of payroll controls

PI-19   Competition, value for money and controls in procurement

PI-20   Effectiveness of internal controls for non-salary expenditure

PI-21   Effectiveness of internal audit

(iii) Accounting, Recording and Reporting

PI-22   Timeliness and regularity of accounts reconciliation

PI-23   Availability of information on resources received by service delivery units

PI-24   Quality and timeliness of in-year budget reports

PI-25   Quality and timeliness of annual financial statements

(iv) External Scrutiny and Audit

PI-26   Scope, nature and follow-up of external audit

PI-27   Legislative scrutiny of the annual budget law

PI-28   Legislative scrutiny of external audit reports

DONOR PRACTICES

D-1   *Predictability of Direct Budget Support*

D-2   *Financial information provided by donors for budgeting and reporting on project and program aid*

D-3   *Proportion of aid that is managed by use of national procedures*

---

[56] See: http://www.pefa.org/sites/pefa.org/files/attachments/PMFEng-finalSZreprint04-12_1.pdf. The estimate of 'PFM Quality' in this paper excludes (a) PI-1 through PI-4 which measure PFM outcomes, (b) indicators PI-13 to PI-15 which cover transparency and effectiveness of tax administration, and (c) D1 to D3, which are the donor-related indicators

References

Alesina, Alberto, Roberto Perotti (1996) Fiscal Discipline and the Budget Process. *American Economic Review, Papers and Proceedings* 86, 401-407

Andrews, M. (2009) Isomorphism and the Limits to African Public Financial Management Reform, *Faculty Research Working Paper RWP 09-012*. Cambridge, MA: Harvard Kennedy School

Andrews, M. (2010) How Far Have Public Financial Management Reforms Come in Africa? *Faculty Research Working Paper RWP 10-018*. Cambridge, MA: Harvard Kennedy School

Auty, R (2000) How Natural Resources Affect Economic Development. *Development Policy Review* 18, 347-364

Baunsgaard, T., M. Villafuerte, M. Poplawski-Riberio, and C. Richmond (2012), *Fiscal Frameworks for Resource Rich Developing Countries*, International Monetary Fund.

Bell, C. (2011) Buying Support and Buying Time: The Effect of Regime Consolidation on Public Goods Provision, *International Studies Quarterly* 55(3).

Belotti, F., S. Daidone, G. Ilardi, V. Atella (2012) Stochastic Frontier Analysis using Stata, *CEIS* 10(12), No. 251.

Bluhm, R. and A. Szirmai (2012) Institutions and long-run growth performance: An analytic literature review of the institutional determinants of economic growth. *UNU-MERIT Working Paper Series on Institutions and Economic Growth: IPD WP02*. Maastricht: UNU-MERIT.

Cruz, C and P. Keefer (2012) *Programmatic Parties and the Politics of Bureaucratic Reform*. Available at: http://147.142.190.246/joomla/peio/files2013/papers/Cruz,%20Keefer%2028.10.2012.pdf.

Daban-Sanchez, T and J.-L. Helis (2013) Public Financial Management in Natural Resource Rich Countries, in: M. Cangiano, T. Curristine, M. Lazare, Public Financial Management and its Emerging Architecture, Washington DC: IMF.

De Renzio, P. (2009) Taking Stock: What Do PEFA Assessments Tell us about PFM Systems across Countries? *ODI Working Paper no. 302*. London: ODI.

de Renzio, P., Gomez, P. and Sheppard, J. (2009) 'Budget Transparency and Development in Resource-Dependent Countries.' International Social Science Journal 57(13): 57-69.

De Renzio, P., M. Andrews, Z. Mills (2011) Does Donor support to Public financial Management reforms in developing countries work? An Analytical Study of Quantitative Cross-Country Evidence. *ODI Working Paper no. 329*. London: ODI.

Farrell, M.J. (1957) The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society* 120(3): 253-282.

Greene, W., (2005) Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis* 23: 7-32.

Lake, David and Matthew Baum, 2001., The Invisible Hand of Democracy: Political Control and the Provision of Public Services. *Comparative Political Studies* 34(6) 587-621.

Moore, M. (2004) Revenues, State Formation, and the Quality of Governance in Developing Countries, *International Political Science Review* 25 (3), 297–319.

Von Hagen, J. and I. Harden (1996) Budget Processes and Commitment to Fiscal Discipline. *IMF Working Paper 96/78*. Washington, DC: IMF.

PEFA Secretariat (2009), Issues in Comparison and Aggregation of PEFA Assessment Results Over Time and Across Countries

PEFA Secretariat (2010) Survey of PEFA Partners' Use of PEFA Assessments for Internal Purposes.

Prakash, T and E. Cabezon (2008) Public Financial Management and Fiscal Outcomes in Sub-Saharan African Heavily-Indebted Poor Countries. *IMF Working Paper no. 08/217*. Washington, DC: IMF.

Pretorius and Pretorius (2008) A Review of the PFM Reform Literature. *Evaluation Working Paper EV 698*. London: DFID.

Prichard, W. and D. Leonard (2010) Does Reliance on Tax Revenue Build state Capacity in Sub-Saharan Africa? *International Review of Administrative Sciences*, 76 (4) [pages].

Tommasi 1999, Managing Government Expenditures, Manila: Asian Development Bank
World Bank (2008) *Governance, Growth, and Development Decision-Making—Reflections by Douglass North, Daron Acemoglu, Francis Fukuyama, and Dani Rodrik.* Washington, DC: World Bank.

Vlaicu, R. Marijn Verhoeven, Francesco Grigoli and Zachary Mills (2014) Multiyear budgets and fiscal performance: Panel data evidence , *Journal of Public Economics* 111 (C): pp. 79-95.

World Bank (2013) *Beyond the Annual Budget. Global Experience with Medium Term Expenditure Frameworks*. Washington DC: World Bank.