

Taking Stock: What do PEFA Assessments tell us about PFM systems across countries?

Paolo de Renzio

Working Paper 302

Results of ODI research presented
in preliminary form for discussion
and critical comment

Working Paper 302

**Taking Stock:
What do PEFA Assessments tell us about
PFM systems across countries?**

Paolo de Renzio

May 2009

Overseas Development Institute
111 Westminster Bridge Road
London SE1 7JD

ISBN 978 0 85003 904 7
Working Paper (Print) ISSN 1759 2909
ODI Working Papers (Online) ISSN 1759 2917

© Overseas Development Institute 2009

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the publishers.

Contents

Author's note	iv
Executive summary	v
1. Introduction	1
2. Methodology	3
3. Summary of results	5
3.1 General characteristics	9
3.2 Economic characteristics	11
3.3 Political characteristics	14
3.4 Basic regression analysis	15
4. Conclusions and next steps	18
Bibliography	20

Author's note

Paolo de Renzio is a Research Associate of the Centre for Aid and Public Expenditure at the Overseas Development Institute, and a doctoral candidate at the Department of Politics and International Relations of the University of Oxford. The research for this paper was carried out in the period August 2007 – April 2008. Access to data by the PEFA Secretariat is gratefully acknowledged. Thanks also to Bill Dorotinsky, Steve Knack, Frans Ronsholt, and Geoff Handley for comments on earlier drafts of this paper. Usual disclaimers apply.

Executive summary

The Public Financial Management (PFM) Performance Measurement Framework, an indicator-based assessment tool developed by the Public Expenditure and Financial Accountability (PEFA) initiative, was launched in 2005 and has been applied so far in over 60 countries. PEFA reports provide detailed accounts of the performance of PFM systems along various dimensions. This paper is based on the results of the 57 PEFA assessments completed as of August 2007. It looks at comparative cross-country PFM performance, overall and across the different budget dimensions defined by the PEFA methodology (out-turns, cross-cutting features, budget cycle), and analyses differences linked to certain country characteristics which might have an influence over PFM system performance, using both bivariate and multivariate analysis. It is based on a numerical conversion of the letter-scores used in the assessments, a methodology which can be considered controversial but which nevertheless yields some interesting results.

Two overall findings become immediately evident. First of all, there is a large variation in overall average scores, ranging from Norway's maximum score of 3.44 (roughly equivalent to a B+) to a minimum score of 1.46 (roughly equivalent to a D+). There are 14 countries that fall below the 2.00 mark (i.e. whose average score is below C), including countries from a range of regions and with different levels of income. The second immediate interesting finding is that average scores tend to deteriorate the further one moves down the various phases of the budget cycle. Given the large number of countries and dimensions involved, overall comparisons are not very useful in terms of detecting specific issues and trends. Instead, the paper compares PEFA scores looking at different country characteristics, including region, population size, and level of income, degree of dependency on foreign aid or natural resources, and strength of democratic institutions.

Taken one by one, these dimensions show some interesting trends, with countries in certain categories showing a better performance than others. However, such binary associations are not necessarily significant from a statistical point of view, and could therefore be potentially misleading. Multivariate regression analysis results highlight that the main factors which are correlated to variations in the overall PEFA score in a statistically significant way are the level of income, country size as measured by the log of the total population, and the degree of aid dependency. Regarding income levels, as expected an increase in per capita income is associated with an increase in overall average PEFA scores. Aid dependency levels are also significant in all models shown, with a positive coefficient but associated with very small changes in PEFA scores, meaning that higher aid dependency levels are associated with marginal improvements in PEFA scores. Finally, as far as population size is concerned, results seem to indicate that larger country size is generally associated with better PFM system performance.

Despite the limitations of the existing data, the analysis highlights some of the interesting comparisons that the existence of such data allows. PEFA assessments are a unique source of information which sheds light on an aspect of governance which until very recently had been mostly overlooked. As more and more assessments are carried out, and repeated in various countries, the availability of more data will allow for a more significant and robust analysis of the determinants and consequences of improvements in PFM system performance, with the potential to generate firmer and more nuanced conclusions. Such analysis could be complemented with a structured comparison of country case studies, in order to allow for a deeper investigation of the large number of factors that are likely to affect the quality of PFM systems and its evolution over time.

1. Introduction

In June 2005, after almost two years of design, a performance monitoring framework for Public Financial Management (PFM) systems was launched by the Public Expenditure and Financial Accountability (PEFA) initiative, a consortium of donor agencies promoting a coordinated approach to the assessment of the quality of PFM systems in developing countries and to the support of related reform programs.

The framework was the result of a long process of consultation and piloting, drawing from a series of previous instruments such as HIPC Assessments, Country Financial Accountability Assessments, various standards and codes promoted by, among others, the IMF and the OECD, and other attempts by donor agencies to measure PFM system performance across countries.

The framework consists of 28 indicators (with an additional three dedicated to donor practices), which attempt to capture the capacity of country PFM systems to deliver positive outcomes along a series of dimensions:

- **Credibility of the budget:** The budget is realistic and is implemented as intended (indicators 1-4);
- **Comprehensiveness and transparency:** The budget and the fiscal risk oversight are comprehensive, and fiscal and budget information is accessible to the public (indicators 5-10);
- **Policy-based budgeting:** The budget is prepared with due regard to government policy (indicators 11-12);
- **Predictability and control in budget execution:** The budget is implemented in an orderly and predictable manner and there are arrangements for the exercise of control and stewardship in the use of public funds (indicators 13-21);
- **Accounting, recording and reporting:** Adequate records and information are produced, maintained and disseminated to meet decision-making control, management and reporting purposes (indicators 22-25); and
- **External scrutiny and audit:** Arrangements for scrutiny of public finances and follow up by executive are operating (indicators 26-28).

Many of the indicators include more specific sub-indicators, and each is scored on a scale from A (highest) to D (lowest), with specific descriptions of the elements necessary to achieve each score for every indicator, and with intermediate scores (denoted by a '+' sign) available for composite indicators that include two or more sub-dimensions.

Since its introduction in 2005, the PEFA framework has been applied in over 60 countries and territories,¹ usually through collaboration between country authorities and donor agencies. The PEFA Secretariat has played an important role in providing feedback and quality assurance on terms of reference and draft reports, ensuring the correct application and adherence to the assessment framework, responding to queries and keeping track of issues and problems arising from the use of the framework.

Of the 57 assessments considered for this paper, 21 were carried out in Sub-Saharan African countries, 14 in Latin America and the Caribbean, 8 in East Asia and the Pacific, 8 in Eastern Europe and Central Asia, 2 in the Middle East and North Africa, 2 in South Asia and 1 in a Western European country (Norway). Of the total, 13 were carried out in 2005, 30 in 2006 and 14 in 2007.² Full details are provided in Table 1 below.

¹ Such as Kosovo and West Bank and Gaza.

² For the few countries where more than one assessment was carried out (Ghana, Malawi, Tanzania and Uganda), this paper utilised the results from the most recent one.

Table 1: Countries included and dates of PEFA assessment (draft or final reports)

Region/country	Assessment date	Region/country	Assessment date
Sub-Saharan Africa		South Asia	
Benin	2007	Afghanistan	2007
Burkina Faso	2006	Bangladesh	2005
Congo, Rep.	2006	Latin America and Caribbean	
Gabon	2005	Barbados	2006
Ghana	2006	Dominica	2007
Guinea Bissau	2006	Dominican Republic	2007
Guinea	2006	Grenada	2006
Kenya	2006	Guatemala	2005
Lesotho	2006	Haiti	2006
Madagascar	2007	Honduras	2005
Malawi	2006	Jamaica	2007
Mali	2006	Nicaragua	2006
Mauritius	2006	Panama	2006
Mozambique	2005	St.Kitts and Nevis	2007
São Tome and Principe	2006	St.Lucia	2006
Sierra Leone	2007	St.Vincent	2006
Swaziland	2006	Trinidad and Tobago	2006
Tanzania	2005	Middle East and North Africa	
Togo	2006	Jordan	2007
Uganda	2007	Syria	2006
Zambia	2005	West Bank and Gaza	2005
East Asia and Pacific		Eastern Europe and Central Asia	
Fiji	2005	Albania	2006
Laos	2006	Armenia	2007
Papua New Guinea	2005	Kosovo	2006
Philippines	2007	Kyrgyz Republic	2005
Samoa	2006	Moldova	2005
Timor Leste	2007	Serbia	2006
Tuvalu	2006	Tajikistan	2006
Vanuatu	2005	Ukraine	2006
		OECD	
		Norway	2007

While most of these assessments utilised the framework in its entirety and were consistent with the methodology, some of them are based on a partial application, either through the coverage of a more limited set of indicators, through the provision of insufficient information to fully justify the scoring, or through the incorrect use of the scoring methodology.³ Furthermore, some of the assessment reports used were still at a draft stage and minor changes in the ratings may be introduced in the final version. Whilst these issues are important for a few of the assessments, they are not believed to undermine the general use of the ratings for comparison on the basis of broad country categories. However, ranking of individual countries on the basis of this data would not be appropriate.

Despite some of these compliance issues, the time seems ripe for a stock-tacking exercise which looks at the results of all existing PEFA assessments from a comparative perspective, using the large quantity of data collected through the country assessments and compiling them in a cross-country database which allows for significant comparisons across countries, groups of countries, regions, budget dimensions, and other potentially interesting categories. While the PEFA framework was not originally intended for the purposes of cross-country comparisons, the purpose of this paper is to show the kinds of results that such analysis is able to generate, and reflect on its usefulness and on the importance of promoting a wider but responsible use of PEFA-generated data.

³ More details on problems with early assessments can be found in PEFA, 2006.

2. Methodology

This paper is based on the results of the 57 PEFA assessments carried out at central government level⁴ and completed before August 2007. It looks at comparative cross-country PFM performance, overall and across the different budget dimensions defined by the PEFA methodology (out-turns, cross-cutting features, budget cycle), and analyses differences linked to certain country characteristics which might have an influence over PFM system performance (such as region, level of income, population, degree of dependency on aid or on resource revenues, administrative heritage, etc.). In order to facilitate comparisons, all PEFA scores have been converted to numerical values, according to the table below.⁵

Table 2: PEFA Score Conversions

PEFA Score	A	B+	B	C+	C	D+	D
Num. Value	4	3.5	3	2.5	2	1.5	1

Numerical values are then averaged across the various budget dimensions and by country, using the scores for the 28 main indicators, and excluding sub-indicators and donor practices. The exclusion of sub-indicators is due to the need to simplify and limit the scope of the analysis at this stage. Donor practices are excluded because the purpose of this paper is to compare country PFM system performance and not donor performance per se, regardless of the impact it might have on country PFM systems. These adaptations also take care of the fact that in some assessments scores for sub-indicators and donor practices were not included. In the database, there are a total of 95 missing scores, out of a total of 1,596 (or about 6%). Some of the indicators worst affected by missing scores are the ones on inter-governmental fiscal relations (PI-8), on tax administration (PI-13, PI-14 and PI-15), on procurement (PI-19) and on budget information at service delivery units (PI-23). It is difficult to assess the impact of missing scores on overall averages, but without any doubt these introduce a potential bias. It could be argued that non-scoring due to lack of information should be included in the rating scale, but this has not been done because it is not certain if non-scoring should be converted to a 'D' level (=1) or even lower (e.g. = 0).

Norway is the only high-level income country where a PEFA assessment has been carried out with the specific purpose of showing how a developed country fares using the methodology. While these results are interesting for general purposes, Norway is excluded from most averages for groups of countries, as its values often distort the results significantly. In some of the comparisons using different country characteristics, some countries also have had to be left out due to lack of data.

The use of average values has already been used in a number of analytical reports issued by donor agencies, but could be criticised on a number of grounds, the main one being that the various dimensions of the PEFA methodology actually measure very different things, which are not necessarily amenable to quantitative conversions, calculations and analysis. Also, the use of averages is based on the assumption that all indicators are equally important. This might be problematic for a number of reasons. For some parts of the framework, for example, some indicators may actually be 'more important' than others, in the sense that a low score may undermine the significance of scores in other indicators. Furthermore, when looking at cross-country comparisons, it is not always possible to ensure that the definitions utilised in different assessments (for example of 'arrears') are compatible, or that the quality and coverage of the assessments is consistent.

⁴ Assessments carried out at local government level were not included.

⁵ The 1-4 scale is of course somewhat arbitrary, but is meant to reflect the fact that a 'D' score in many cases denotes a deficient system, not a non-existent one.

While these observations are obviously valid,⁶ recognising these potential limitations should not prevent an attempt at exploring the possible uses of such numerical conversions, and testing whether they generate interesting and/or counter-intuitive findings. Moreover, the rigorous design of the methodology, and its clear definition of the thresholds for scoring the indicators, ensures that in most cases cross-country comparisons are indeed feasible, with possible measurement errors and biases being compensated by the large number of countries included in the sample.

⁶ The PEFA Secretariat is in fact drafting a paper on the pros and cons of aggregating scores for cross-country comparisons.

3. Summary of results

Table 3 presents the summary results.⁷ Two overall findings become immediately evident. First of all, there is a large variation in overall average scores, ranging from Norway's maximum score of 3.44 (roughly equivalent to a B+) to a minimum score of 1.46 (roughly equivalent to a D+). Interestingly, however, there are countries whose performance is quite close to the top one, with scores of 3.23 and 3.21. There are also 14 countries that fall below the 2.00 mark (i.e. whose average score is below C), including countries from a range of regions and with different levels of income.

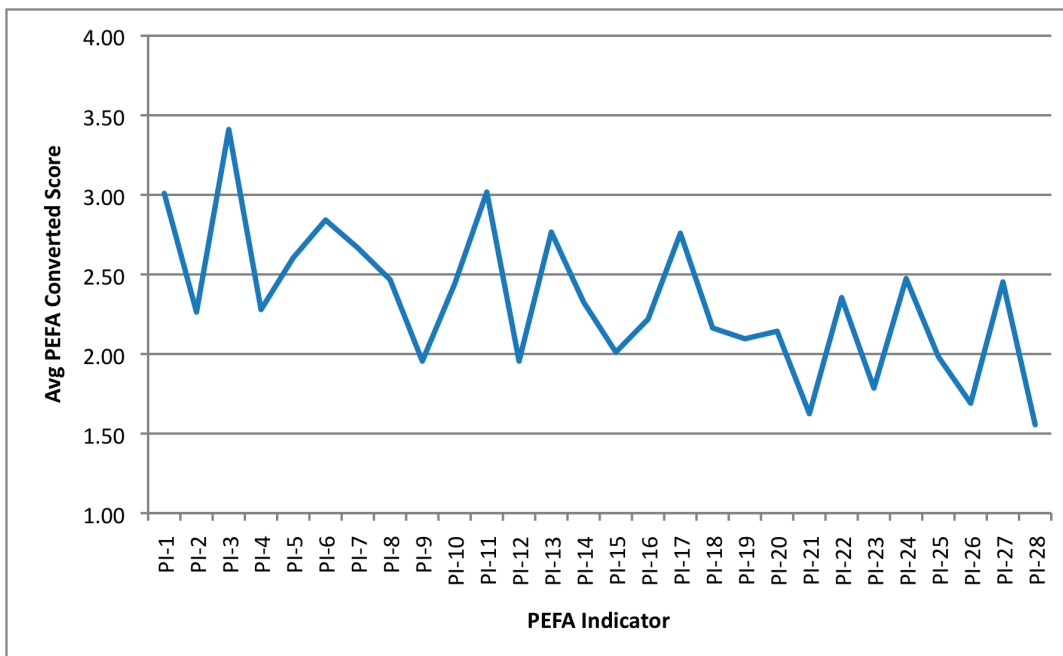
Table 3: Summary results by budget dimension and indicator

Budget Dimension and Indicators		Average	St. Dev.	Min	Max
Budget Credibility					
PI-1	Aggregate expenditure out-turn compared to original approved budget	3.01	1.00	1	4
PI-2	Composition of expenditure out-turn compared to original approved	2.26	1.10	1	4
PI-3	Aggregate revenue out-turn compared to original approved budget	3.41	0.95	1	4
PI-4	Stock and monitoring of expenditure payment arrears	2.28	1.14	1	4
Average		2.74	0.64	1	4
Comprehensiveness & Transparency					
PI-5	Classification of the budget	2.61	0.90	1	4
PI-6	Comprehensiveness of information included in budget documentation	2.84	0.84	1	4
PI-7	Extent of unreported government operations	2.67	0.96	1	4
PI-8	Transparency of inter-governmental fiscal relations	2.47	0.94	1	4
PI-9	Oversight of aggregate fiscal risk from other public sector entities.	1.96	0.84	1	4
PI-10	Public access to key fiscal information	2.44	0.80	1	4
Average		2.50	0.55	1.5	3.67
Policy-Based Budget					
PI-11	Orderliness and participation in the annual budget process	3.02	0.71	1.5	4
PI-12	Multi-year perspective in fiscal planning, expenditure policy and budgeting	1.96	0.76	1	3.5
Average		2.49	0.63	1.25	4
Control in Execution					
PI-13	Transparency of taxpayer obligations and liabilities	2.77	0.74	1.5	4
PI-14	Effectiveness of measures for taxpayer registration and tax assessment	2.33	0.63	1	4
PI-15	Effectiveness in collection of tax payments	2.01	0.97	1	4
PI-16	Predictability in the availability of funds for commitment of expenditures	2.22	0.95	1	4
PI-17	Recording and management of cash balances, debt and guarantees	2.76	0.75	1	4
PI-18	Effectiveness of payroll controls	2.16	0.93	1	4
PI-19	Competition, value for money and controls in procurement	2.10	0.70	1	3.5
PI-20	Effectiveness of internal controls for non-salary expenditure	2.14	0.73	1	4
PI-21	Effectiveness of internal audit	1.63	0.60	1	3.5
Average		2.23	0.49	1.5	3.44
Accounting & Reporting					
PI-22	Timeliness and regularity of accounts reconciliation	2.35	0.96	1	4
PI-23	Availability of information on resources received by service delivery units	1.79	0.96	1	4
PI-24	Quality and timeliness of in-year budget reports	2.47	0.88	1	4
PI-25	Quality and timeliness of annual financial statements	1.98	0.96	1	4
Average		2.15	0.67	1	3.88
External Scrutiny					
PI-26	Scope, nature and follow-up of external audit	1.69	0.74	1	3.5
PI-27	Legislative scrutiny of the annual budget law	2.45	0.86	1	4
PI-28	Legislative scrutiny of external audit reports	1.56	0.65	1	3.5
Average		1.90	0.55	1.00	3.17
Overall Average		2.33	0.44	1.46	3.44

⁷ Results for each country are not shown as many of the reports were not publicly available at the time of finalising this paper.

The second immediate interesting finding is that average scores tend to deteriorate the further one moves down the various phases of the budget cycle, from formulation to execution, reporting and scrutiny. On average, even though again this hides a lot of variation (average standard deviations by budget dimension range from 0.49 to 0.67), the countries under observation fare quite well with regard to general issues of *budget credibility* (2.74) and *comprehensiveness and transparency* (2.50), and with regard to *policy-based budgeting* (2.49), which looks at the initial stages of the budget process. However, their performance gradually deteriorates when one looks at *predictability and control in budget execution* (2.23), *accounting, recording and reporting* (2.15), and *external scrutiny and audit* (1.90).⁸ Figure 1 below shows this trend indicator by indicator. There are various possible explanations for this. Certainly, many donor programs have focused more on ‘upstream’ budget formulation (e.g. macro-fiscal frameworks, Medium Term Expenditure Frameworks, budget classification, etc.). Moreover, reforms in budget execution, often linked to the implementation of integrated financial management systems (IFMS), take longer to implement and are often fraught with failures.⁹ On the other hand, while issues of scarce capacity and donor focus may have played a role, other factors could also have prevented ‘downstream’ systems from being strengthened, possibly linked to the stronger political and bureaucratic resistance that they are likely to encounter.

Figure 1: Average PEFA converted scores by indicator



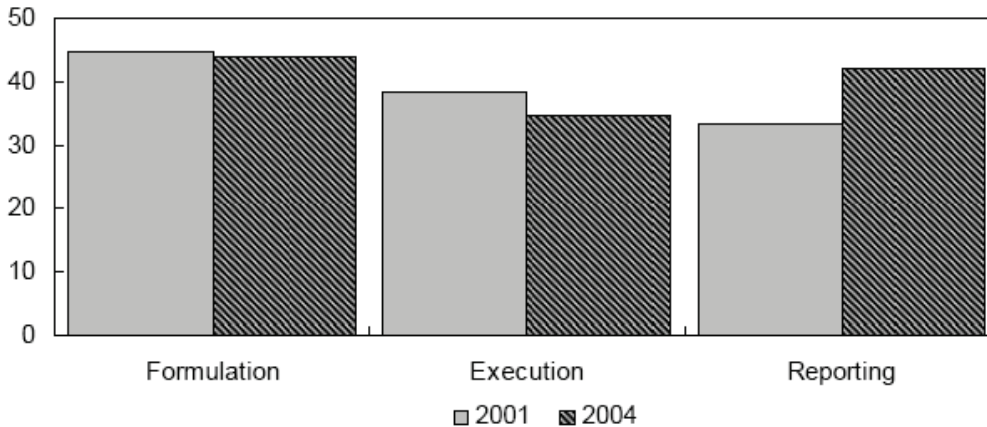
In fact, this seems to support the opinion that, over time, many developing countries have achieved better results in the area of budget formulation than in the more difficult areas of budget execution, monitoring and audit. Figure 2, taken from the World Bank/IMF 2005 Board Paper summarising the results of HIPC Assessments,¹⁰ shows at least partially (the result is more evident for 2001 than for 2004) how countries score better on budget formulation than they do on execution and reporting. At the same time, this finding calls into question the usefulness of having well-formulated and comprehensive budgets, when they are executed with inadequate controls, insufficient reporting and limited scrutiny by parliaments and audit institutions.

⁸ Computing Spearman rank correlation indices for the six budget dimensions shows that the differences in averages are not heavily distorted by rank and are mostly statistically significant.

⁹ For example, see Diamond and Khemani (2005).

¹⁰ IDA/IMF (2005).

Figure 2: Percent of benchmarks met (by PFM category)



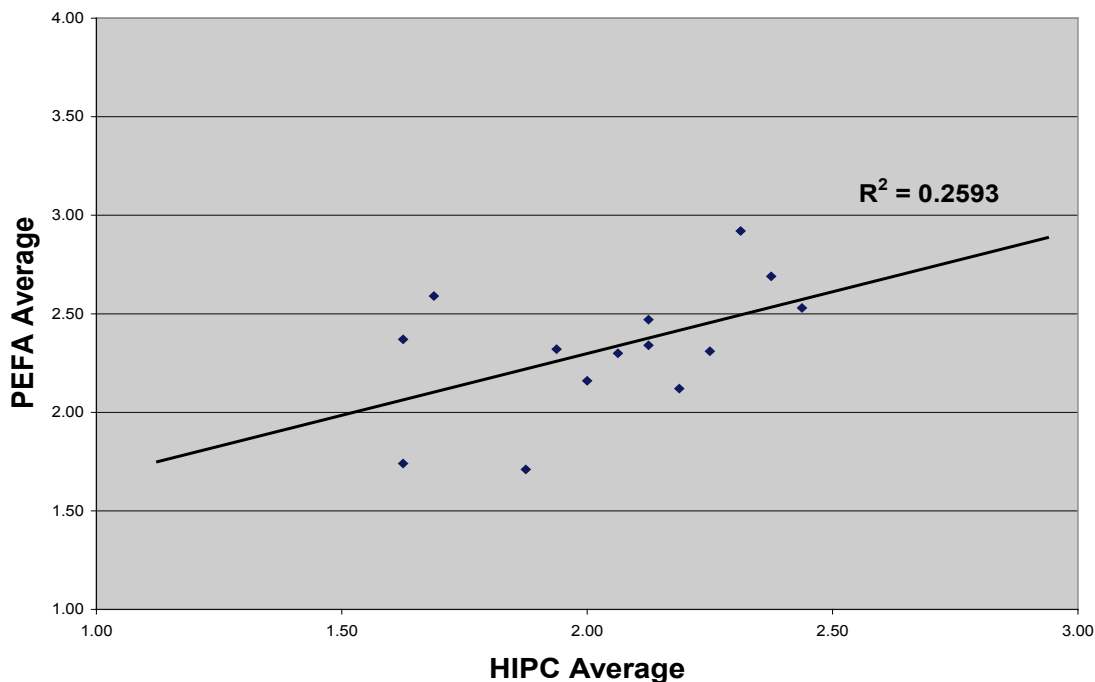
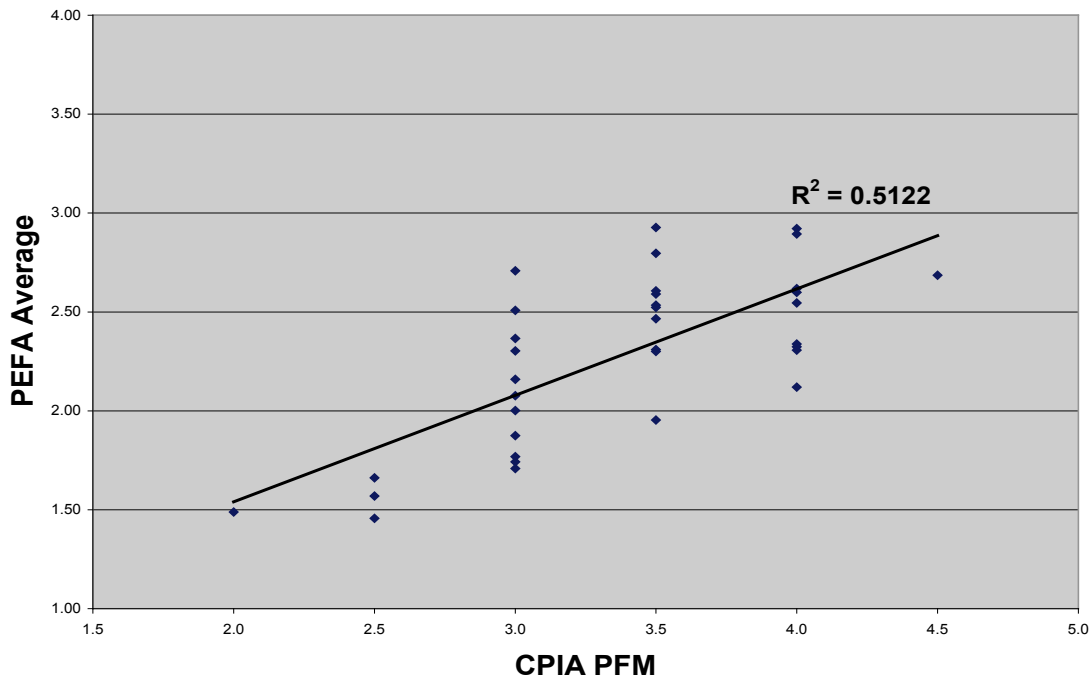
Source: IDA/IMF (2005).

It is also interesting to note, in Table 3, how within each budget dimension, averages for different indicators can vary substantially. As a consequence, for example, the higher averages for budget credibility are driven mostly by aggregate out-turns, and for policy-based budgeting by the regular working of the budget process. Lower averages in the 'downstream' phases of the budget cycle, on the other hand, are substantially worsened by ineffective internal and external auditing processes.

There are few ways of checking whether the average scores generated through PEFA assessments accurately describe the differences that exist among different countries. A simple test of that is to see how the average PEFA scores correlate with other similar measures of the quality of budget systems. The Country Policy and Institutional Assessments (CPIA) produced by the World Bank include a PFM indicator. The correlation between PEFA overall average scores and the CPIA PFM scores is quite high (0.72), but this is likely to be quite a spurious comparison, as the results of PEFA assessments are increasingly being used in the compiling of CPIA indicators. Another possible comparable index is that produced for the HIPC AAP exercise (IDA/IMF, 2005). Using a numerical conversion system similar to the one used in this paper for the PEFA scores,¹¹ the correlation between HIPC indicators and PEFA ones is slightly lower (0.51), and is based on the small sample of countries in which both assessments took place. Figure 3 shows the fit between the sets of indicators.

¹¹ See de Renzio and Dorotinsky (2007).

Figure 3: Comparing PEFA averages with other PFM indicators



Unfortunately, there are no other available cross-country indicators that look at the quality of PFM systems in a broad range of countries and with a comprehensive approach. Two existing indices address more specific aspects of PFM systems. The Open Budget Index (www.openbudgetindex.org) focuses on the issue of budget transparency (its correlation with indicator PI-10 is 0.26), and in its release for 2006 covered 59 countries. The Global Integrity Index (www.globalintegrity.org) mostly focuses on issues of budget accountability (its correlation with indicator PI-27 is 0.11), and in its latest release for 2007 covered 76 countries. Other governance indicators, such as the one on 'bureaucratic quality' produced by the International Country Risk Guide (ICRG) and the one on 'government effectiveness' compiled by the World Bank Institute, have a much wider coverage but actually measure something quite different from PEFA. Moreover, these broader indicators, especially the WBI ones

which are obtained aggregating a number of other independently-produced indicators, have been criticised on the basis that they suffer from a number of measurement biases and problems.¹²

Given the large number of countries and dimensions involved, overall comparisons are not very useful in terms of detecting specific issues and trends. If one is interested in exploring and understanding some of the factors that can explain differences in PFM system performance across countries, classifying countries according to various characteristics, and grouping them together can possibly lead to more interesting comparisons. For example, countries can be grouped according to some general characteristics (region, population size, and administrative heritage), economic characteristics (level of income, degree of dependency on foreign aid or natural resources) and political characteristics (democracy level, political system and degree of press freedom). In the rest of this section, therefore, two approaches are offered to highlight more specific findings on the factors influencing the quality of PFM systems across countries. First, some simple bivariate statistics are shown for selected variables, to see if any of these are associated with higher or lower levels of PFM system performance, both overall and by budget dimension. Secondly, in the final part of the section, the preliminary results of a multivariate regression analysis are also presented, to see in which ways the various factors interact with each other and are jointly associated with changes in overall PFM system performance. It is very important to stress that any association between certain country characteristics and better PFM system performance cannot be interpreted as a causal relation, in whichever sense. Rather, it simply reflects a statistical correlation, an association of concurrent evidence that could be explained in a number of different ways.

3.1 General characteristics

3.1.1 Regional groups

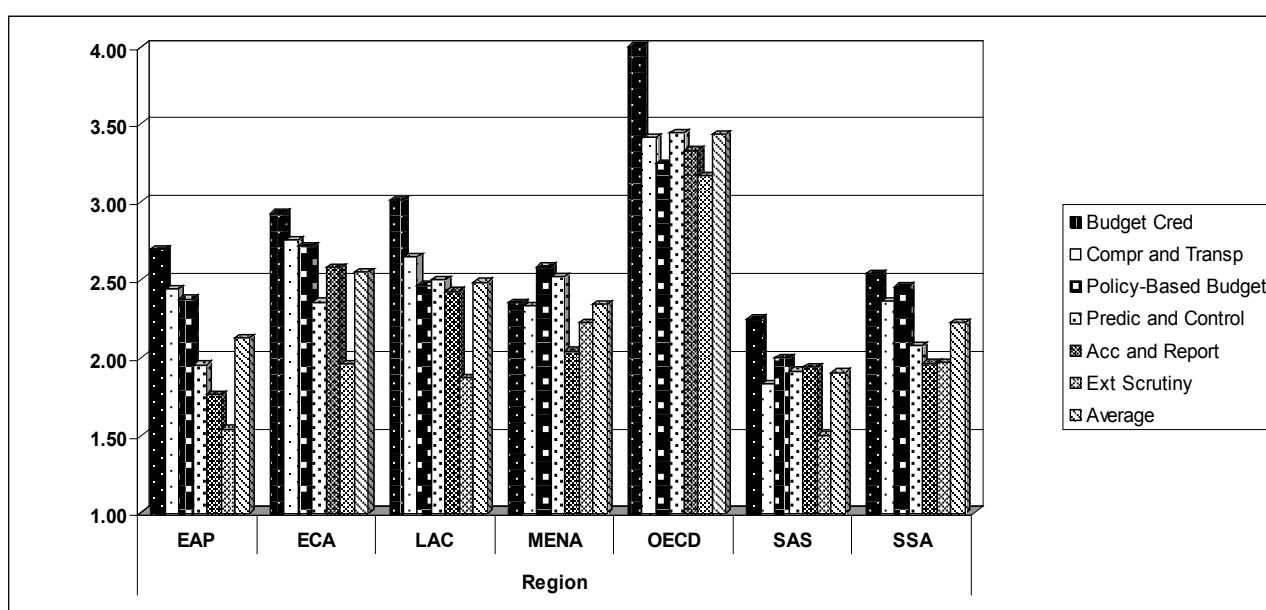
One factor that might be associated with changes in PFM system performance is the geographical location of a country, or the region to which it belongs, which in turn can reflect certain historical and cultural factors.

Table 4 and Figure 4 summarise the results grouping countries by region. In general terms, and disregarding the only OECD country (Norway), Eastern European (2.55) and Latin American (2.49) countries perform better, reaching an average roughly equivalent to a C+ score, while South Asian countries (although the sample is not really representative, as it includes only two countries) average 1.91 (slightly below a C). The differences among different budget dimensions, confirming the declining trend across the various phases of the budget cycle, are highest for East Asian and Pacific and Latin American countries, while they are less accentuated in the Middle East and North Africa, where in fact *policy-based budgeting* and *predictability and control in budget execution* score highest (again, these results need to be taken carefully, as the sample from the MENA region also only includes two countries), or in the Eastern Europe and Central Asia region, which score relatively better on *accounting, recording and reporting*.

¹² See, for example, Arndt and Oman (2006) and Kaufmann and Kraay (2008).

Table 4: Averages by region and budget dimension¹³

	Region						
	EAP (8)	ECA (8)	LAC (14)	MENA (3)	OECD (1)	SAS (2)	SSA (21)
Budget Credibility	2.70	2.93	3.01	2.35	4.00	2.25	2.54
Comprehensiveness and Transparency	2.44	2.75	2.65	2.33	3.42	1.83	2.36
Policy-Based Budgeting	2.38	2.72	2.46	2.58	3.25	2.00	2.45
Predictability and Control in Execution	1.96	2.36	2.50	2.52	3.44	1.92	2.08
Accounting and Reporting	1.76	2.58	2.43	2.04	3.33	1.94	1.96
External Scrutiny	1.54	1.96	1.87	2.22	3.17	1.50	1.97
Average	2.13	2.55	2.49	2.34	3.44	1.91	2.23

Figure 4: Averages by region and budget dimension

3.1.2 Population size

Could the size of a country also be associated with the quality of PFM systems? There could be economies of scale in investing in budget systems in larger countries, or otherwise public finances might be easier to manage in smaller countries.

Figure 5 finds a low negative relation between country size and the overall PEFA average. Table 5 presents the more detailed averages having grouped countries according to the tercile to which they belong if ranked by population size. The evidence here is not in any sense clear, although smaller countries do score slightly better, both in terms of overall averages and for most budget dimensions. The variation, however, is quite minimal, as all groups of countries hover around average scores roughly equivalent to C–C+.

¹³ EAP=East Asia & Pacific, ECA=Eastern Europe & Central Asia, LAC=Latin America & Caribbean, MENA=Middle East & Northern Africa, OECD=Organization for Economic Cooperation and Development, SAS=South Asia, SSA=Sub-Saharan Africa.

Figure 5: PEFA averages and population size

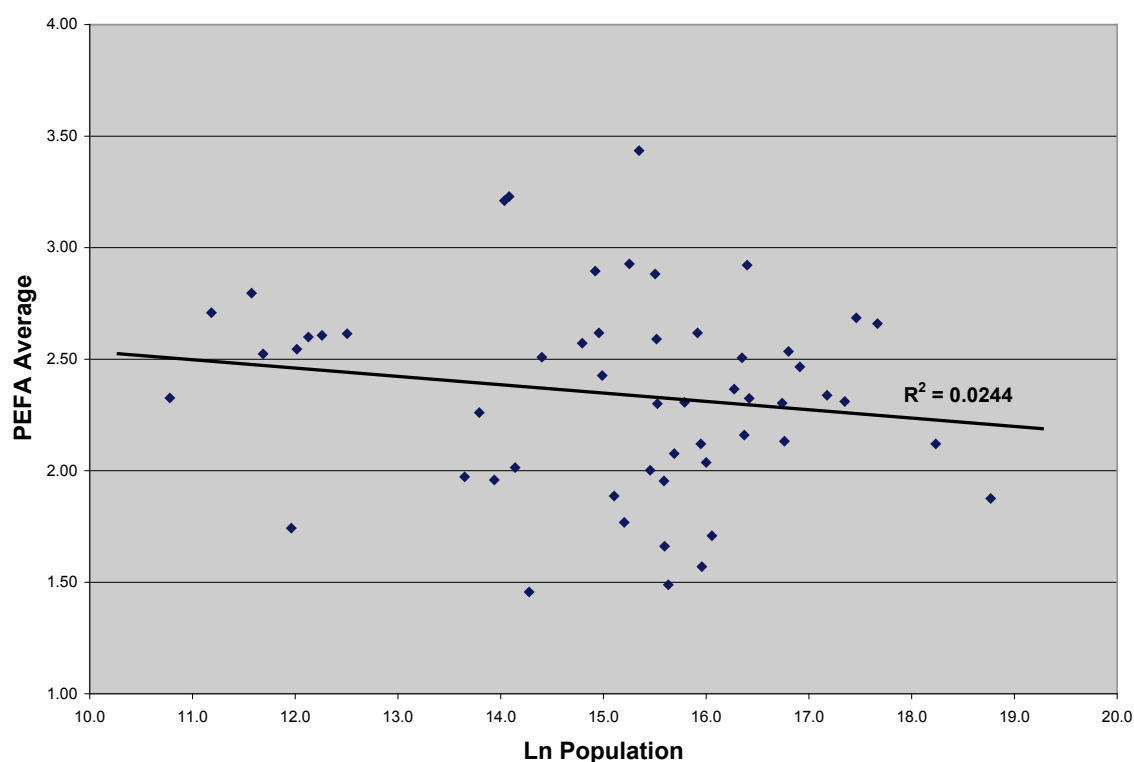


Table 5: Averages by population size and budget dimension

	Population		
	Low	Medium	High
Budget Credibility	2.99	2.59	2.63
Comprehensiveness and Transparency	2.62	2.45	2.43
Policy-Based Budgeting	2.57	2.49	2.44
Predictability and Control in Execution	2.33	2.25	2.09
Accounting and Reporting	2.19	2.07	2.17
External Scrutiny	1.87	1.81	1.97
Average	2.43	2.28	2.29

3.2 Economic characteristics

3.2.3 Income level

Another obvious factor that could influence (or be influenced by) the quality of PFM systems is the income level that each country enjoys. Figure 6, quite expectedly, shows a positive correlation between income per capita and PEFA averages. Table 6 and Figure 7 show the results by grouping countries according to World Bank country classification of per capita income.¹⁴ What the data confirms is the hypothesis that countries with a higher level of income perform better in terms of the quality of their

¹⁴ Countries are divided according to 2006 GNI per capita, calculated using the World Bank Atlas method. The groups are: low income (LICs), \$905 or less; lower middle income (LMICs) \$906 – \$3,595; and upper middle income (UMICs) \$3,596 - \$11,115. Norway, with a GNI per capita of \$60,890 was excluded.

PFM systems, both in terms of overall average and for each budget dimension, excluding *policy-based budgeting*, on which upper-middle income countries actually average lower than poorer countries. Overall, while the small sample of upper middle income countries (UMICs) covered, mostly made up of very small countries, come close to a B score (2.80), low income countries are closer to a C score (2.13). Interestingly, in UMICs the weakness of latter phases of the budget cycle seems to be less accentuated, with both *accounting and reporting* and *external scrutiny* reaching much better averages than lower income countries, and comparable to averages for the rest of the budget cycle. This might be related to the growing focus on ‘checks and balances’ that is associated with higher levels of per capita income, or more simply with the fact that richer countries have more resources to devote to improving PFM systems.

Figure 6: PEFA averages and income per capita

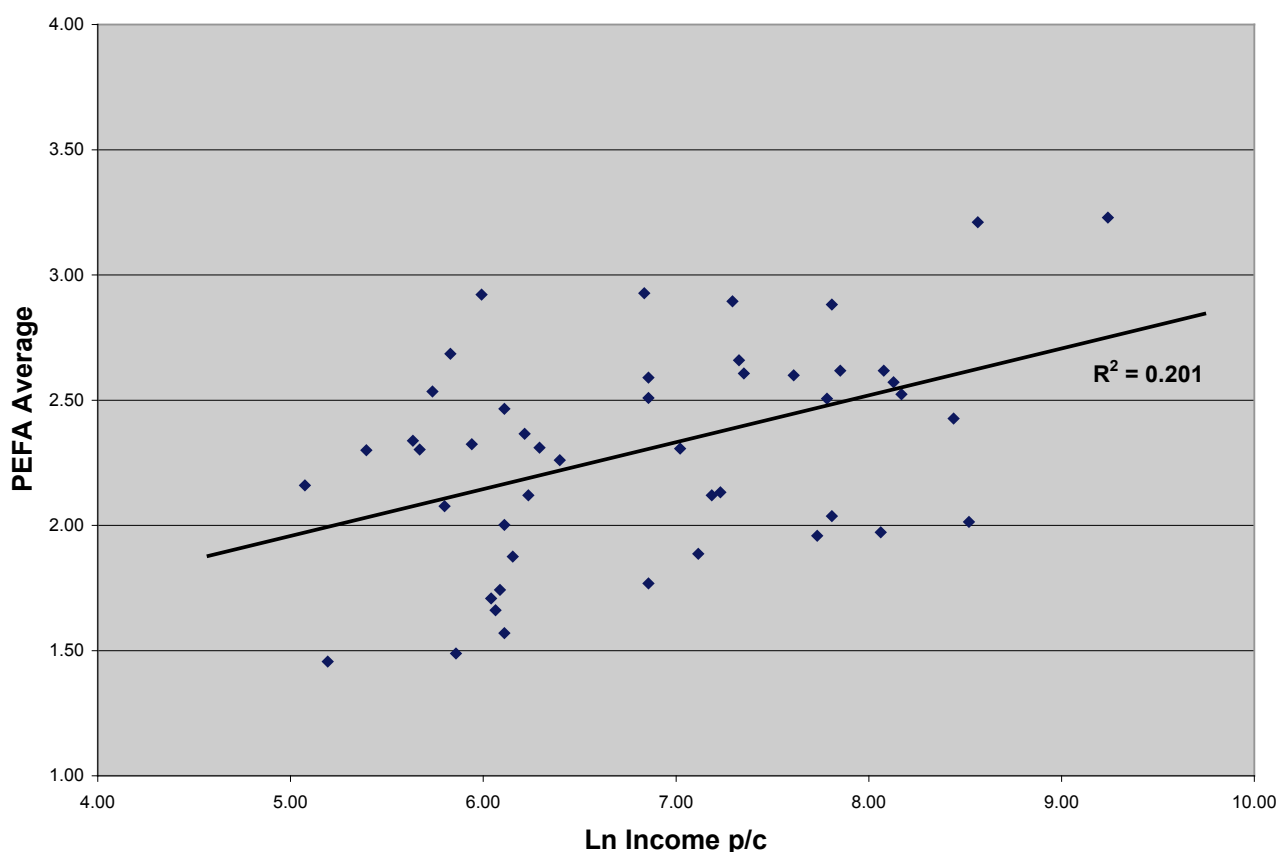
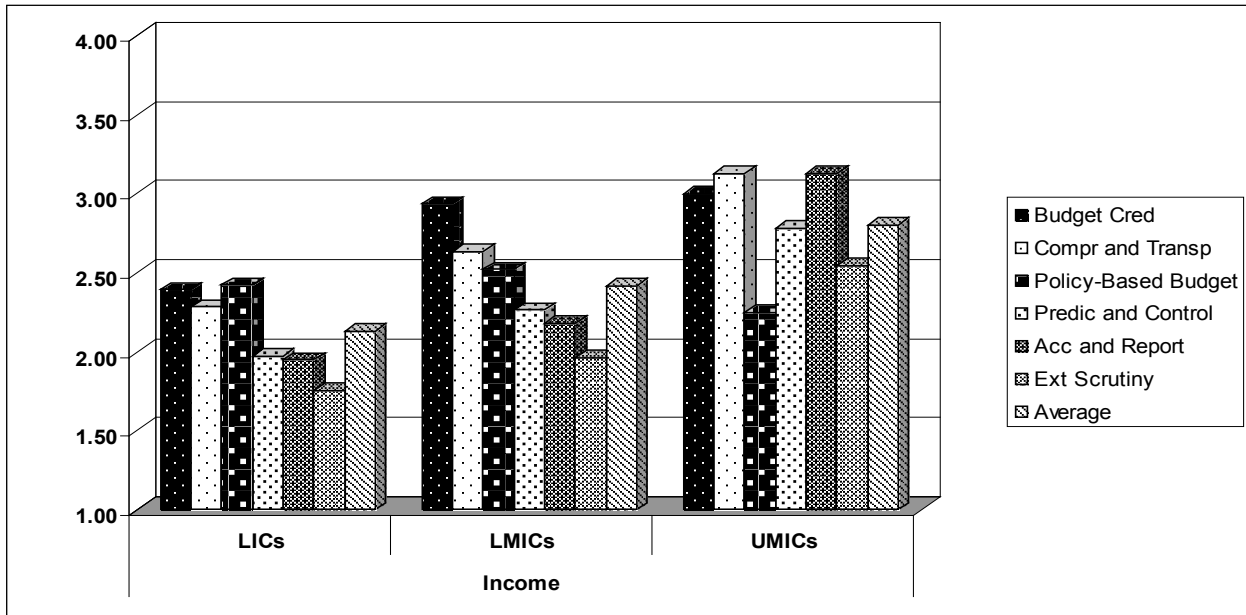


Table 6: Averages by income group and budget dimension

	Income		
	LICs	LMICs	UMICs
Budget Credibility	2.39	2.93	3.00
Comprehensiveness and Transparency	2.28	2.63	3.13
Policy-Based Budgeting	2.42	2.52	2.25
Predictability and Control in Execution	1.97	2.26	2.78
Accounting and Reporting	1.95	2.18	3.13
External Scrutiny	1.75	1.96	2.54
Average	2.13	2.41	2.80

Figure 7: Averages by income group and budget dimension



3.2.4 Aid and resource dependency

Dependency on certain sources of revenue, for example in the form of rents from natural resource extraction or from foreign aid, has been considered as a factor linked to governance more in general, and therefore potentially to changes in the quality of PFM systems.¹⁵ Figure 8, for example, shows the weak negative correlation that exists between levels of aid dependency and PEFA average scores. In Table 7, countries have been grouped according to their degree of dependency on such sources of revenue.¹⁶ What the results show is that resource dependency does not seem to be associated with any significant difference in the quality and performance of PFM systems. For aid dependency, there is a limited improvement in overall average scores associated with lower levels of dependency, but one which is not consistent across budget dimensions. Interestingly, if average scores for donor practices (such as the predictability of disbursements and the percentage of aid which uses national procedures) are included in the analysis, these tend to be lower (1.71) in highly aid dependent countries than in countries that receive little foreign aid (2.05). This might be linked to the degree of fragmentation, overlap and inefficiency that could be associated with a larger donor presence in high aid-dependent countries, and with the subsequent lower ability of aid-dependent governments to enforce discipline among donor agencies.

¹⁵ See, for example, Brautigam (2000), Brautigam and Knack (2004), and Eifert et al. (2003).

¹⁶ For aid dependency, total Official Development Assistance (ODA) as a percentage of GNI in 2004 was used as the indicator. High aid dependency is defined when the indicator is higher than 10%. Medium when it is between 5% and 10%, and low when it is lower than 5%. For resource dependency, see the list contained in IMF (2007:62-3).

Figure 8: PEFA averages and aid dependency

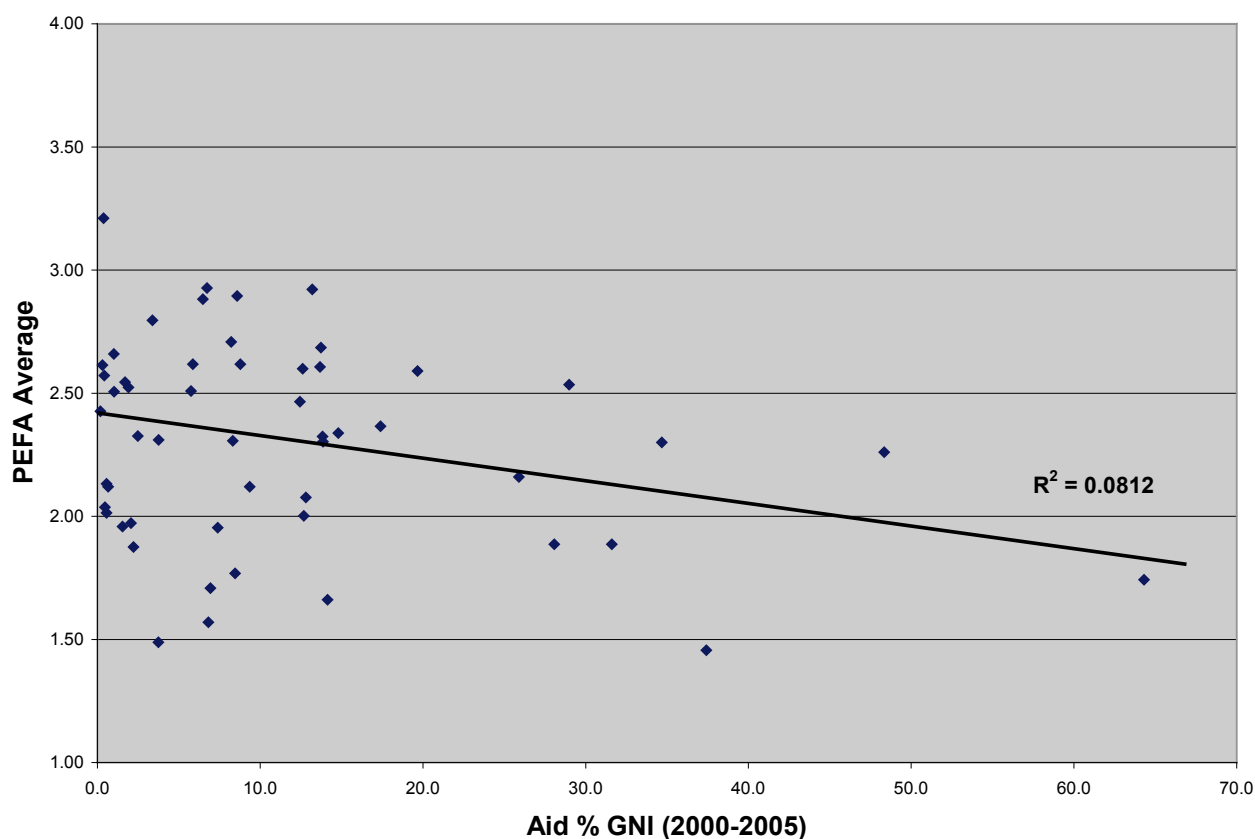


Table 7: Averages by degree of aid/resource dependency and budget dimension

	Aid Dependency			Res Dependency	
	Low	Medium	High	Yes	No
Budget Credibility	2.76	2.76	2.71	2.54	2.75
Comprehensiveness and Transparency	2.57	2.48	2.40	2.51	2.47
Policy-Based Budgeting	2.46	2.62	2.50	2.41	2.48
Predictability and Control in Execution	2.39	2.11	2.07	2.28	2.20
Accounting and Reporting	2.24	1.91	2.14	1.92	2.17
External Scrutiny	1.96	1.71	1.88	2.03	1.86
Average	2.40	2.27	2.28	2.28	2.32
Donor Practices	2.05	1.76	1.71		

3.3 Political characteristics

3.3.5 Democracy level

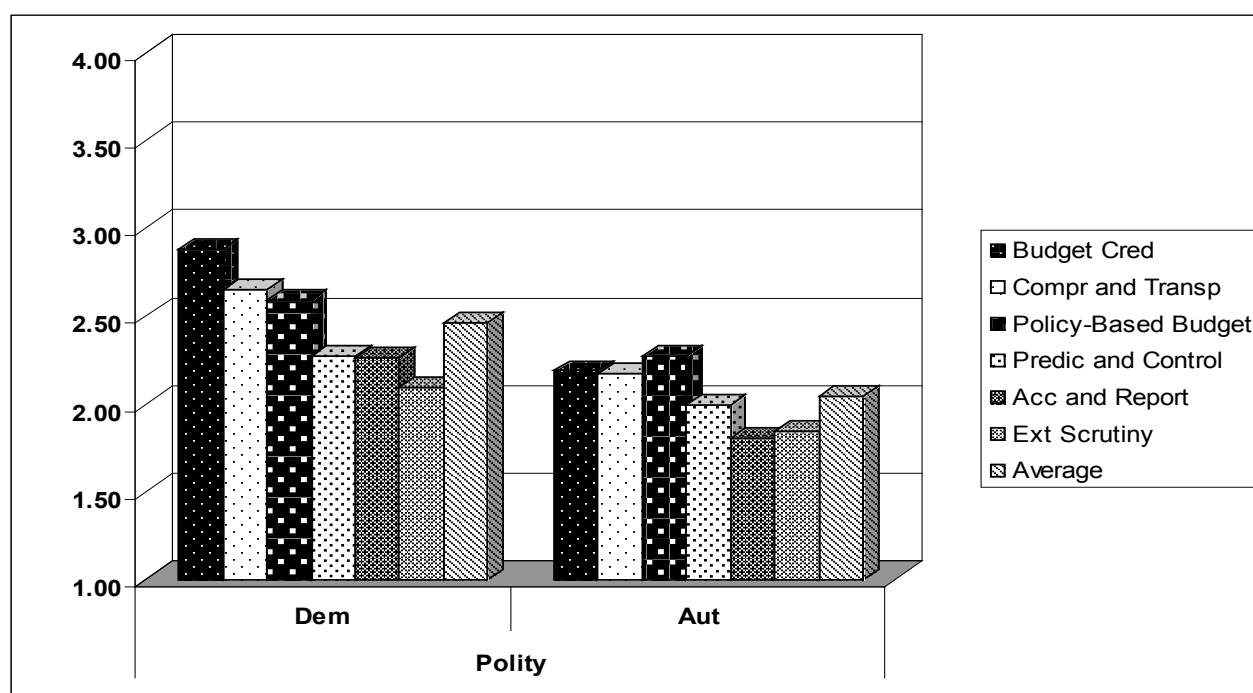
It might also be interesting to investigate whether differences in the nature of political systems are associated with changes in PFM system performance. For this purpose, countries in Table 8 and Figure 9 have been grouped according to how they rank in terms of the openness of their political institutions,

and whether they tend towards a democratic or an autocratic system.¹⁷ Both overall and across all six PFM dimensions, countries with more democratic political institutions rank better (2.46) than countries which tend towards autocracy (2.04).

Table 8: Averages by democracy level and budget dimension

	Polity	
	Dem	Aut
Budget Credibility	2.88	2.19
Comprehensiveness and Transparency	2.65	2.17
Policy-Based Budgeting	2.59	2.27
Predictability and Control in Execution	2.27	1.99
Accounting and Reporting	2.26	1.81
External Scrutiny	2.09	1.84
Average	2.46	2.04

Figure 9: Averages by democracy level and budget dimension



3.4 Basic regression analysis

In previous parts of this section, some interesting insights came from grouping countries according to various categories, and seeing if different groups scored differently on PFM system performance, both overall and by budget dimension. However, these binary associations are not necessarily significant from a statistical point of view, as for each country the overall average PEFA score might be determined

¹⁷ The indicator used is the Combined Polity Score from the Polity IV database (<http://www.cidcm.umd.edu/polity>). Countries scoring from -10 to 0 have been labeled 'autocratic', while those with a score between 1 and 10 have been labeled 'democratic'.

by a number of these factors and categories acting contemporaneously. For example, Latin American countries also tend to have a higher per capita income and therefore be less aid dependent, while.

Multivariate regression analysis is meant to address exactly this issue, looking at how various factors (the explanatory or independent variables) are associated with variation in the main variable of interest (the dependent variable), in this case the overall average PEFA score. Table 9 provides some summary statistics for the main variables included in the regression analysis. These go beyond the ones considered in the bivariate analysis above, to include other relevant variables, namely the type of political system in the various countries (i.e. whether it is presidential, parliamentary, or mixed), the degree of press freedom, and a dummy variable for the administrative heritage of former colonies.

Table 9: Summary statistics

Variable	Obs	Mean	Std Dev	Min	Max
PEFA Score	57	2.329649	.4420446	1.46	3.44
Log GNI	48	6.930603	1.181228	5.075174	11.01682
Log Population	53	15.18106	1.744988	11.51293	18.76993
Aid/GNI (av. 2000-05)	54	11.3037	13.00978	0	64.3
Democracy Level ¹⁸	42	3.833333	5.467941	-9	10
Political System ¹⁹	49	0.6326531	0.9058577	0	2
Press Freedom ²⁰	54	44.90741	19.72286	10	84

As can be seen, the per capita income and population variables were converted to their logarithmic form, given the large range of initial values. In addition to the variables described above, dummy variables for resource dependency, for the various regions and for the different colonial powers were also included in the regression analysis. The estimation method utilised was ordinary least squares (OLS) regression, given the simple nature of the analysis, which is based on a purely cross-sectional database.

Regression results are shown in Table 10, using the PEFA overall average score as the dependent variable, and different combinations of independent variables to explore different possible interactions, rather than test any specific hypotheses.²¹ The results highlight that the main factors which are correlated to variations in the overall PEFA score in a statistically significant way are the level of income, country size as measured by the log of the total population, and the degree of aid dependency. Resource dependency, political variables (including press freedom) and administrative heritage are almost never statistically significant under the various specifications attempted. Regional differences are also mostly not significant.

¹⁸ The 'democracy level' variable is based on the Polity IV 'polity combined score' which ranks countries on a scale from -10 to 10 (more autocratic to more democratic). See <http://www.cidcm.umd.edu/polity>.

¹⁹ The 'political system' variable classifies countries in the basis of whether they are presidential (0), mixed (1) or parliamentary (2). See Beck et al. (2001).

²⁰ Drawn from the Freedom House Freedom of the Press 2007 Survey (see <http://www.freedomhouse.org>). The higher the score, the lower the degree of press freedom in each country.

²¹ Regional and administrative heritage dummies are not shown in detail.

Table 10: OLS regression results

	(1)	(2)	(3)	(4)	(5)
Log per capita income	0.64***	0.39***	0.42***	0.30***	0.28***
Log population	0.28***	0.19**	0.15**	0.11	0.07
Aid dependency	0.05***	0.02**	0.03***	0.12*	0.02*
Resource dependency	-0.34				
Presidential system	-0.36	-0.23	-0.15	0.25*	
Parliamentary system	0.07	0.006	0.33	0.32	
Press freedom	-0.006	-0.006		-0.003	
Democracy Level	-0.0006	0.01		0.01	0.03***
Intercept	-6.11**	-3.05*	-3.08**	-1.74	-0.99
<i>Regional Fixed Effects</i>	<i>YES</i>	<i>YES</i>	<i>YES</i>	<i>NO</i>	<i>NO</i>
<i>Admin. Heritage Fixed Effects</i>	<i>YES</i>	<i>NO</i>	<i>NO</i>	<i>NO</i>	<i>NO</i>
<i>Number of Observations</i>	<i>38</i>	<i>38</i>	<i>43</i>	<i>38</i>	<i>40</i>
<i>Adjusted R²</i>	<i>0.78</i>	<i>0.65</i>	<i>0.49</i>	<i>0.46</i>	<i>0.41</i>

Note: * significant at 10%, ** at 5%, *** at 1%. All regional and administrative heritage dummies never achieve statistical significance.

Regarding **income levels**, as expected an increase in per capita income is associated with an increase in overall average PEFA scores. More specifically, a doubling of the income level is associated with a notable increase in the PEFA score, between 0.28 and 0.64, depending on the specification and holding all other variables constant. This corresponds to almost half a score point (i.e. a shift from C to C+) on the PEFA scale.

Aid dependency levels are also significant in all models shown, with a positive coefficient which is associated with very small changes in PEFA scores, meaning that higher aid dependency levels are associated with marginal improvements in PEFA scores. This seems to counter the results of other research which sees higher aid dependency levels generally associated with lower governance scores.²² On the other hand, it could be related to the fact that countries that receive more aid, especially if in the form of budget support, are likely to receive more assistance related to PFM systems performance and reform, or to causation in the opposite direction, where countries with better PFM systems are more likely to receive additional donor assistance. However, the size of the coefficients is very small, meaning that only large increases in the level of aid dependency would be associated with significant improvements in average PEFA scores. In most cases, in order to see a positive half score point change, aid dependency would need to increase by 10-20 percentage points with relation to gross national income.

Finally, as far as **population size** is concerned, results seem to indicate that larger country size is generally associated with better PFM system performance, contrary to what shown in Figure 4 and Table 3 above, where a bivariate relation showed smaller countries scoring better. In this case, a doubling of the population would be associated with an increase in the average PEFA score between 0.15 and 0.28, again depending on the specification and holding all other variables constant.

While regression analysis can show which factors better explain the variation in the overall PEFA scores, there are a number of serious limitations to this kind of analysis, mostly due to the nature of the data. First of all, while PEFA data are available for 57 countries, the lack of a full dataset for the other variables means that the total number of observations included in each regression hovers around 40, and therefore limits the scope of the analysis. Secondly, the analysis has treated the numerical PEFA scores as a cardinal variable, while it is not obvious that this assumption is fully justifiable. More importantly, however, the cross-sectional nature of the data, where all countries are compared for a single time-period, does not allow for the inclusion of any analysis of country-specific trends.

²² Brautigam and Knack (2004).

4. Conclusions and next steps

This paper has looked at the results of 57 PEFA assessments which have been carried out until August 2007 around the world. The purpose of the analysis was to verify the usefulness of broad comparisons across countries using numerically-converted PEFA scores, both in terms of overall averages and of averages for the main budget dimensions covered by the framework. In order to do this, average scores were compared across countries grouped according to various categories, such as region, population, income level, dependency on aid and resource revenues, and democratic institutions. Taken one by one, these dimensions show some interesting trends, with countries in certain categories (e.g. higher income, lower aid dependency, more democratic institutions) showing a better performance than others. However, such binary associations are not necessarily significant from a statistical point of view, and could therefore be potentially misleading.

Multivariate regression analysis shows that in fact only some factors reach statistical significance in explaining levels of overall PEFA scores. Among these, income level is the most significant one. As countries' income levels rise, the performance of their PFM systems improves substantially. That these two phenomena are closely associated is not surprising. The endogenous nature of such association, however, possibly prevents any significant policy lessons. While it might be tempting to state that as a consequence a lot of effort should be put in improving PFM systems, given their effect on income levels, it is much more likely to be the case that causation goes the other way around. The rise in income levels may generate both the necessity and the pressure for government to improve the management of public resources, while at the same time increasing existing capacity to back and sustain such improvements.

Aid dependency levels are also significantly and positively associated with PEFA scores. Again, the direction of influence cannot be determined within the present analysis. On one hand, higher levels of aid dependency may mean that donors invest more in PFM reform programs, which in turn have a positive effect on PFM system performance. On the other hand, the relationship may be generated by the fact that countries with relatively better PFM systems attract more aid, and therefore appear as more aid dependent. Moreover, even if causality could be established, the small size of the coefficients means that massive injections of additional aid would be needed for relatively small improvements in PFM systems.

Despite the limitations of the existing data, which covers a limited set of countries for a single period of time, the analysis highlights some of the interesting comparisons that the existence of such data allows. PEFA assessments are a unique source of information, a new dataset which sheds light on an aspect of governance which until very recently had been mostly overlooked. As more and more assessments are carried out, and repeated in various countries, the availability of more data will allow for a more significant and robust analysis of the determinants and consequences of improvements in PFM system performance, with the potential to generate much more interesting findings and recommendations. The basic analysis adopted in this paper could also be improved and extended, using more sophisticated and adequate statistical techniques, and better or more disaggregated explanatory variables. In turn, quantitative analysis could be complemented by structured comparisons of country-specific case studies that can delve deeper into the large number of context-specific variables that might affect the quality of PFM systems in different countries.

It would therefore be very useful for all PEFA assessments not to simply remain as stand-alone documents which feed into country-level policy dialogue and reform design, but for the indicator information and scores to be coded in a comprehensive database which includes a series of other data on each country, such as the categories which have been included in this analysis. This could prove a very useful comparative tool for the donors involved in the PEFA consortium, and would allow the PEFA Secretariat to respond in a much more effective way to various queries and to generate analytical reports that highlight in-country and cross-country trends and comparisons. These could include a more detailed comparative analysis for specific indicators, or the identification of case studies for more

in-depth research, for example by focusing on outlier countries with specific characteristics and a particularly positive or negative PFM system performance.

Bibliography

- Arndt, C. and Oman, C. (2006) *Uses and Abuses of Governance Indicators*. Paris: OECD, and.
- Beck, Clarke, G., Groff, A., Keefer, P., and P. Walsh (2001) 'New Tools in Comparative Political Economy: The Database of Political Institutions', *World Bank Economic Review* 15 (1):165-176.
- Brautigam, D. (2000) *Aid Dependence and Governance*. Stockholm: Almqvist & Wicksell.
- Brautigam, D. and Knack, S. (2004) 'Foreign Aid, Institutions, and Governance in Sub-Saharan Africa'. *Economic Development and Cultural Change* 52 (2):255-285.
- de Renzio, P. and Dorotinsky, W. (2007) *Tracking Progress in the Quality of PFM Systems in HIPC*s. Washington, DC: PEFA Secretariat. (<http://www.pefa.org/Documents.htm#trackprogress>)
- Diamond, J. and Khemani, P. (2005) *Introducing Financial Management Information Systems in Developing Countries*. IMF Working Paper WP/05/196. Washington DC: International Monetary Fund. (<http://www.imf.org/external/pubs/ft/wp/2005/wp05196.pdf>)
- Eifert, B., Gelb, A. and Tallroth, N. B. (2003) 'Managing Oil Wealth', *Finance and Development*, 40. Washington, DC: International Monetary Fund.
- IDA/IMF (2005) *Update on the Assessments and Implementation of Action Plans to Strengthen Capacity of HIPC*s to Track Poverty-Reducing Public Spending. Washington, DC: International Development Association and International Monetary Fund. (<http://www1.worldbank.org/publicsector/pe/HIPC/HIPCBoardPaperApril2005.pdf>)
- IMF (2007) *Guide on Resource Revenue Transparency*. Washington, DC: International Monetary Fund.
- Kaufmann, D. and Kraay, A. (2008) 'Governance Indicators: Where Are We, Where Should We Be Going?', *Policy Research Working Paper 4370*. Washington, DC: World Bank.
- PEFA (2006) *Report on Early Experience from Application of the Framework* published in November 2006 (www.pefa.org). Washington, DC: PEFA Secretariat