Budget Credibility: What Can We Learn from PEFA Reports? | July 2018

The ability to implement a budget is a crucial component of a sound public financial management (PFM) system.[1] However, budget credibility is an area that is often neglected and under-researched. Despite substantial investments to improve PFM systems and strengthen capacities, many countries still struggle to implement their budgets as originally planned. This paper seeks to provide a global analysis of budget credibility based upon the data available from the PEFA framework.[2]

WHAT IS BUDGET CREDIBILITY?

Budget credibility describes the ability of governments to ***accurately*** and ***consistently*** meet their expenditure and revenue targets. A government can be accurate but not consistent, consistent but not accurate, neither, or both. These concepts also apply at sub-aggregate levels, where there may be substantially more variation.

The PEFA 2010 Framework defines budget credibility (or "Credibility of the Budget") as "The budget is realistic and is implemented as intended."[3] It scores different dimensions of budget credibility through the following four indicators:

TABLE 1. PEFA CREDIBILITY OF THE BUDGET PERFORMANCE INDICATORS

PI-1	Aggregate expenditure outturn compared to original approved budget
PI-2	Composition of expenditure outturn compared to original approved budget
PI-3	Aggregate revenue outturn compared to original approved budget
PI-4	Stock and monitoring of expenditure payment arrears[4]

[1] This report was drafted by Zac Mills and edited by IBP.

[2] The PEFA framework is a universally accepted standardized tool to assess a country's PFM performance. The framework was updated in 2016 and PEFA now refers to budget credibility as "budget reliability."

[3] PEFA framework, page 2.

[4] Analysis of PI-4, a broader issue than budget credibility alone, is omitted from this study.

The deviation between planned and actual spending is measured over a 12-month period (the budget year) for each of the previous three years.

Deviations in expenditure and revenue may have important implications for macroeconomic stability, public service delivery, and social welfare. If expenditure is under-executed, beneficiaries may not receive crucial services. Similarly, if expenditure is over-executed without a corresponding match in revenues, there may be an increase in the budget deficit, debt service financing, public debt levels, and other stresses to macroeconomic stability.

In a given year, there are nine different possible scenarios for a government's expenditure and revenue outturns (see Table 2). The middle scenario of ***Planned / Planned*** is the optimal outcome, but achieving this outcome depends on a number of important factors, including the robustness of forecasts, changes in government policy, and unanticipated events.

TABLE 2. POSSIBLE SCENARIOS FOR EXPENDITURE AND REVENUE DEVIATIONS

		Revenue		
		Under	*Planned*	*Over*
Expenditure	**Under**	Under / Under	Under / Planned	Under / Over
	Planned	Planned / Under	Planned / Planned	Planned / Over
	Over	Over / Under	Over / Planned	Over / Over

Note: Expenditure is first, followed by Revenue

DESCRIPTION OF PEFA DATA AND DESCRIPTIVE STATISTICS

The data are drawn from PEFA indicators PI-1, PI-2, and PI-3. Recently, the PEFA Secretariat introduced the 2016 PEFA assessment methodology, which updated and improved the 2011 PEFA assessment methodology (key differences are summarized in Annex 1). In order to ensure valid comparisons between the 2011 and 2016 methodologies, the raw data underlying each performance indicator score was compiled and the new dimensions in the 2016 methodology were not considered. Therefore, indicators, PI-1, PI-2i, and PI-3i are the main indicators used in the analysis. While there are some minor differences in the scoring methodologies for these indicators between the 2011 and 2016 methodologies, analyzing the raw data maintains compatibility and increases the sample size. Using the most recent publicly available PEFA report for each country since 2012, the raw expenditure and revenue outturn data were compiled for 74 countries. The summary statistics are provided in Table 3.

TABLE 3. PEFA SUMMARY STATISTICS FOR PI-1, PI-2I, AND PI-3I

Variable	Obs	Mean	Std. Dev.	Min	Max
pi1t3	74	100.5	14.5	64.2	170
pi1t2	74	99.5	11.8	67	131.2
pi1t1	74	100.0	14.1	67	135.1
pi2it3	71	16.0	10.0	1.8	49.5
pi2it2	71	14.8	10.0	2.3	49.5
pi2it1	71	15.9	10.7	1.7	58.1
pi3it3	74	101.3	16.6	71.9	194.7
pi3it2	74	101.0	13.7	64	151.3
pi3it1	74	99.8	11.6	74.4	128.1

A breakdown of PEFA assessments by region and the corresponding outturn and deviation averages are presented in Table 4. Sub-Saharan Africa (AFR) is the most represented region with 30 PEFA assessments (41 percent of the total), following by the Europe and Central Asia (ECA) region with 13, and the East Asia and Pacific (EAP) region and the Latin America and Caribbean (LAC) region with 12 each.[5] For the whole sample, the outturns averaged a perfect 100.0 percent for PI-1 and a near perfect average of 100.8 percent for PI-3i. Across regions, average outturns do not differ significantly from the global averages. The exception is the LAC region, which has a significantly lower score in the deviation of expenditure composition.

TABLE 4. BUDGET CREDIBILITY BY REGION

Region	#	PI-1 Average	PI-2i Average[6]	PI-3i Average
AFR	30	101.5	19.0	101.9
ECA	13	100.5	12.2	100
EAP	12	97.9	17.5	102.9
LAC	12	99.0	9.8	96.3
SAR	5	96.6	14.7	99.5
MENA	2	102.4	15.6	106.3
Total	***74***	***100.0***	***15.6***	***100.8***

The same analysis categorized by income level does yield significantly different results (Table 5). The high income sample is distorted by the small sample size, but the upper middle income, lower middle income, and low income countries had on average similar outturns for PI-1 and PI-3. The key area of difference was in PI-2i, where, on

[5] The South Asian region (SAR) had five and the Middle East and North Africa (MENA) region had two assessments in the sample.
[6] Burundi and Ghana were dropped from all the PI-2i analysis since this indicator was not rated in their PEFA assessments.

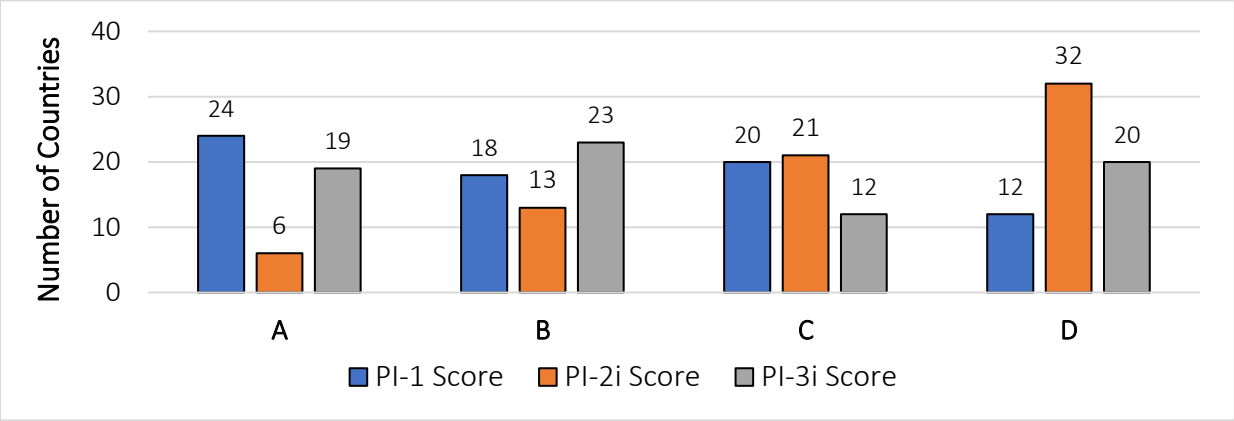
average, there is a clear expansion in the deviation of expenditure composition along the continuum from high income to low income countries.

TABLE 5. BUDGET CLASSIFICATION BY INCOME CLASSIFICATION

Income Level	#	PI-1 Average	PI-2i Average	PI-3i Average
High income	2[7]	100.7	12.2	82.5
Upper middle income	18	98.5	12.5	101.1
Lower middle income	34	100.4	15.3	101.5
Low income	20	100.7	19.4	101.2
Total	***74***	***100.0***	***15.6***	***100.8***

A further analysis by PEFA score shows that a similar number of countries scored an "A" or "B" on PI-1 and PI-3i, but there is a striking linear trend on PI-2i where most countries scored a "D" (see Figure 1). This initial finding suggests that, on average, budget credibility is less an aggregate issue and more of a compositional issue.

Figure 1. PEFA SCORES FOR PI-1, PI-2I, AND PI-3I



WHICH COUNTRIES OVERSPEND / UNDERSPEND?

The aggregated groupings and averages mask the considerable variation at the country level. Table 6 illustrates that during the three-year period of assessment, there were 30 countries that on average underspent (defined as budget execution less than 97.5 percent), 18 countries that on average executed their budget as planned (defined as budget execution between 97.5 percent and 102.5 percent), and 26 countries that on average overspent their budget (defined as budget execution greater than 102.5 percent). These findings, based on a different definition

[7] The two high income countries are Capo Verde and Antigua and Barbuda.

than the PEFA framework, suggest that the accuracy element of budget credibility faces widespread challenges at the aggregate level. [8]

A more granular analysis of PI-1 shows that country averages also mask the volatility within countries. Figure 2 provides the three data points, the corresponding average and standard deviation for each country, and is sorted from the lowest to. [9] This analysis illustrates a different narrative – although countries may overspend or underspend on average, there is a significant divergence in the consistency of expenditure outturns across the three-year period. For example, South Africa, El Salvador, and the Kyrgyz Republic were the most consistent performers, and also had a near perfect outturn of 100. High standard deviations are likely due to some combination of shocks and challenges in forecasting methods. However, there were also several countries with a low standard deviation that consistently deviated from the planned outcome, such as Capo Verde, Rwanda, and Vietnam (over) and Kosovo (under), which suggests a prevailing incentive to under or over-estimate expenditure.[10]

[8] We use all three years of raw data, rather than ignoring one of the years as PEFA does in calculating its scores. In addition, using the raw data circumvents the minor differences in scoring methodologies between 2011 and 2016.

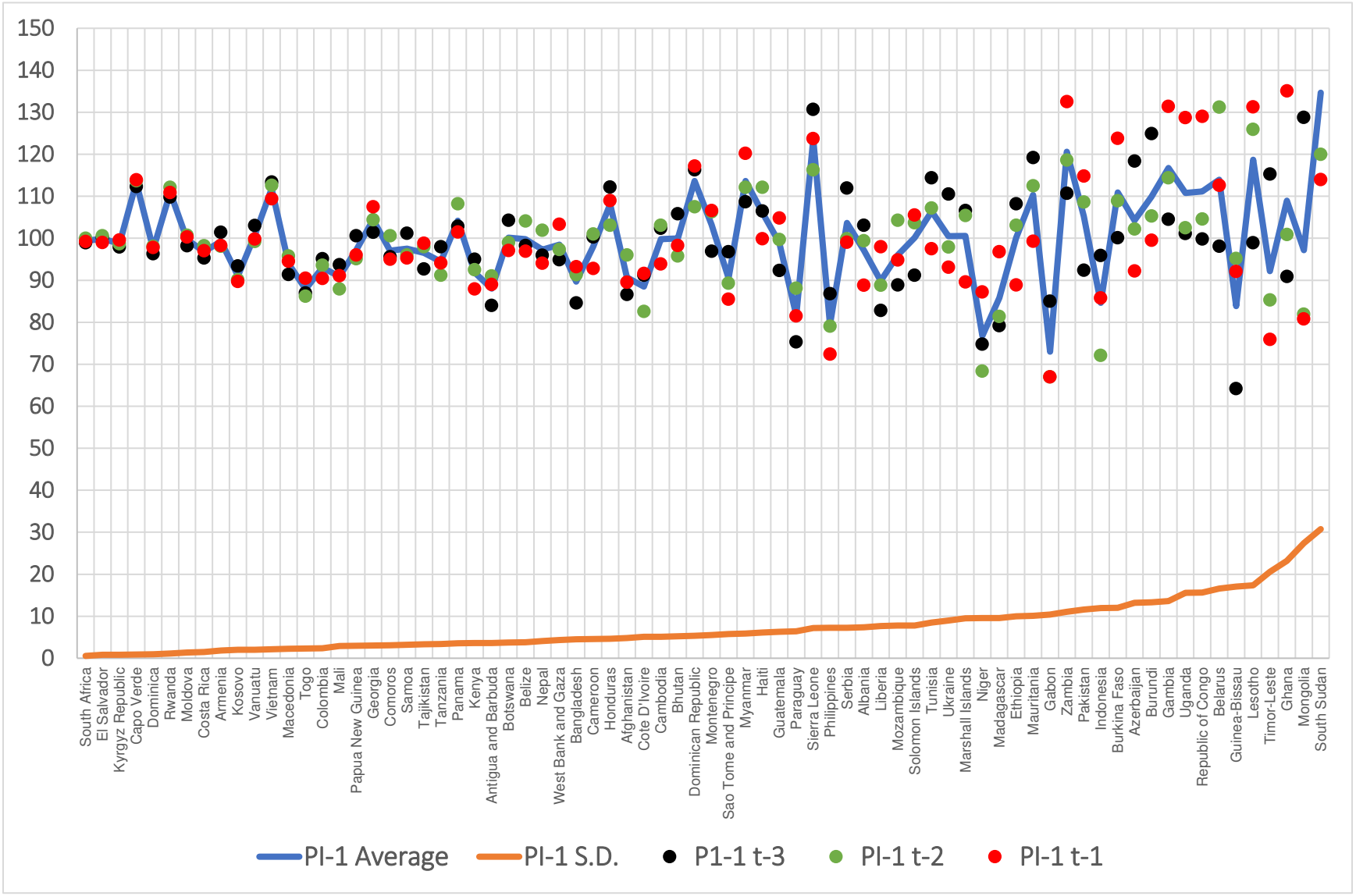
[9] The formula for standard deviation is $\sigma = \sqrt{\frac{\Sigma(X-\bar{X})^2}{N-1}}$

[10] For example, Ethiopia had budget outturns of 108.2 (t-3), 103.1 (t-2), and 88.9 (t-1), with a mean of 100.1. Calculating the standard deviation requires subtracting the deviation from each mean, squaring the difference, and adding the total together (66.2 + 9.2 + 124.7) which is equal to 200. This total is then divided by the number of observations (3) minus one (so 2), which gives 100, and then taking the square root gives 10. In contrast, South Africa had much more consistent budget outturns with a standard deviation of 0.74.

TABLE 6. LIST OF COUNTRIES OVER-SPENDING VS. UNDER-SPENDING (AVERAGE OF PI-1, DERIVED FROM T-3, T-2, T-1)

Country	Under-spend	Country	As planned	Country	Over-spend
Gabon	73.0	Samoa	97.5	Montenegro	103.3
Niger	76.8	Cameroon	98.0	Serbia	103.6
Philippines	79.4	West Bank and Gaza	98.5	Panama	104.2
Paraguay	81.6	Kyrgyz Republic	98.7	Azerbaijan	104.3
Guinea-Bissau	83.8	Guatemala	98.9	Georgia	104.4
Indonesia	84.6	Armenia	99.3	Pakistan	105.3
Madagascar	85.8	South Africa	99.4	Haiti	106.2
Togo	87.9	Moldova	99.7	Tunisia	106.4
Antigua and Barbuda	88.0	Belize	99.8	Honduras	108.1
Cote D'Ivoire	88.5	Cambodia	99.8	Ghana	109.0
Bangladesh	89.7	El Salvador	99.9	Burundi	109.9
Liberia	89.9	Bhutan	100.0	Mauritania	110.3
Sao Tome and Principe	90.5	Botswana	100.1	Uganda	110.8
Afghanistan	90.7	Ethiopia	100.1	Burkina Faso	110.9
Mali	90.9	Solomon Islands	100.1	Rwanda	110.9
Kosovo	91.1	Marshall Islands	100.5	Republic of Congo	111.1
Kenya	91.8	Ukraine	100.5	Vietnam	111.8
Timor-Leste	92.2	Vanuatu	100.7	Capo Verde	113.3
Colombia	93.0			Dominican Republic	113.7
Macedonia	93.9			Myanmar	113.7
Tanzania	94.4			Belarus	114.0
Mozambique	96.0			Gambia	116.8
Tajikistan	96.5			Lesotho	118.7
Costa Rica	96.8			Zambia	120.6
Albania	97.1			Sierra Leone	123.6
Comoros	97.1			South Sudan	134.7
Mongolia	97.2				
Papua New Guinea	97.2				
Nepal	97.3				
Dominica	97.4				

Figure 2. PI-1 EXPENDITURE OUTTURN DATA POINTS, AVERAGE, AND STANDARD DEVIATION



Note: South Sudan had its t-3 observation at 170, which extends beyond the chart.

A regional breakdown shows that the AFR region has the highest percentage of countries with high PI-1 standard deviation, with 18 out of 30 countries having a standard deviation greater than 7, followed by the EAP and ECA regions (see Table 7). In contrast, only nine out of 74 countries had a PI-1 standard deviation of less than two, with five countries having a standard deviation less than one. These findings indicate that only a small sample of countries are consistent in their aggregate expenditure outturn from year to year.

TABLE 7. REGIONAL COMPARISON OF PI-1 STANDARD DEVIATION

Region	0-1	1-2	2-3	3-5	5-7	7-10	>10
AFR	2	1	2	5	2	5	13
EAP			3	1	2	3	3
ECA	1	2	2	2	1	3	2
LAC	2	1	1	4	4		
MENA				1		1	
SAR				3	1		1
Total	***5***	***4***	***8***	***16***	***10***	***12***	***19***

IS BUDGET CREDIBILITY MORE CHALLENGING AT THE COMPOSITIONAL LEVEL?

Yes. Figure 3 illustrates that few countries have a low average deviation score for PI-2i, the measure of compositional credibility at ministry level. In this Figure, countries are sorted by the PI-2i average score to show that only 18 of 71 countries had an average score less than 10, which is roughly the cutoff to qualify for a "B" grade in the PEFA assessment. Conversely, there were 32 countries that had an average score greater than 15 (roughly the cutoff for a "D" grade). Interestingly, the average standard deviation is less for PI-2i than PI-1, which indicates that many countries consistently score poorly on this measure.

To investigate this question further, the complete ministry-level outturn for 28 countries was compiled from the P-2 annexes in each PEFA report. Using this raw data, a new deviation index was created that weights the share of each ministry in the overall budget.[11] Figure 4 confirms that the average budget deviation at the compositional level is greater than the average aggregate budget deviation for all countries.[12]

[11] In this index, the deviation from the approved budget for each ministry in every year was computed and the absolute value was taken. Next, this absolute deviation was multiplied by the percentage share of the approved ministry budget in the total approved budget, and then the weighted absolute deviations totals were summed for all ministries to provide an overall compositional budget deviation index for each country in each year.

[12] The aggregate budget deviation is simply the absolute value of the aggregate budget outturn minus 100.

Figure 3. PI-2I COMPOSITIONAL EXPENDITURE OUTTURN DATA POINTS, AVERAGE, AND STANDARD DEVIATION

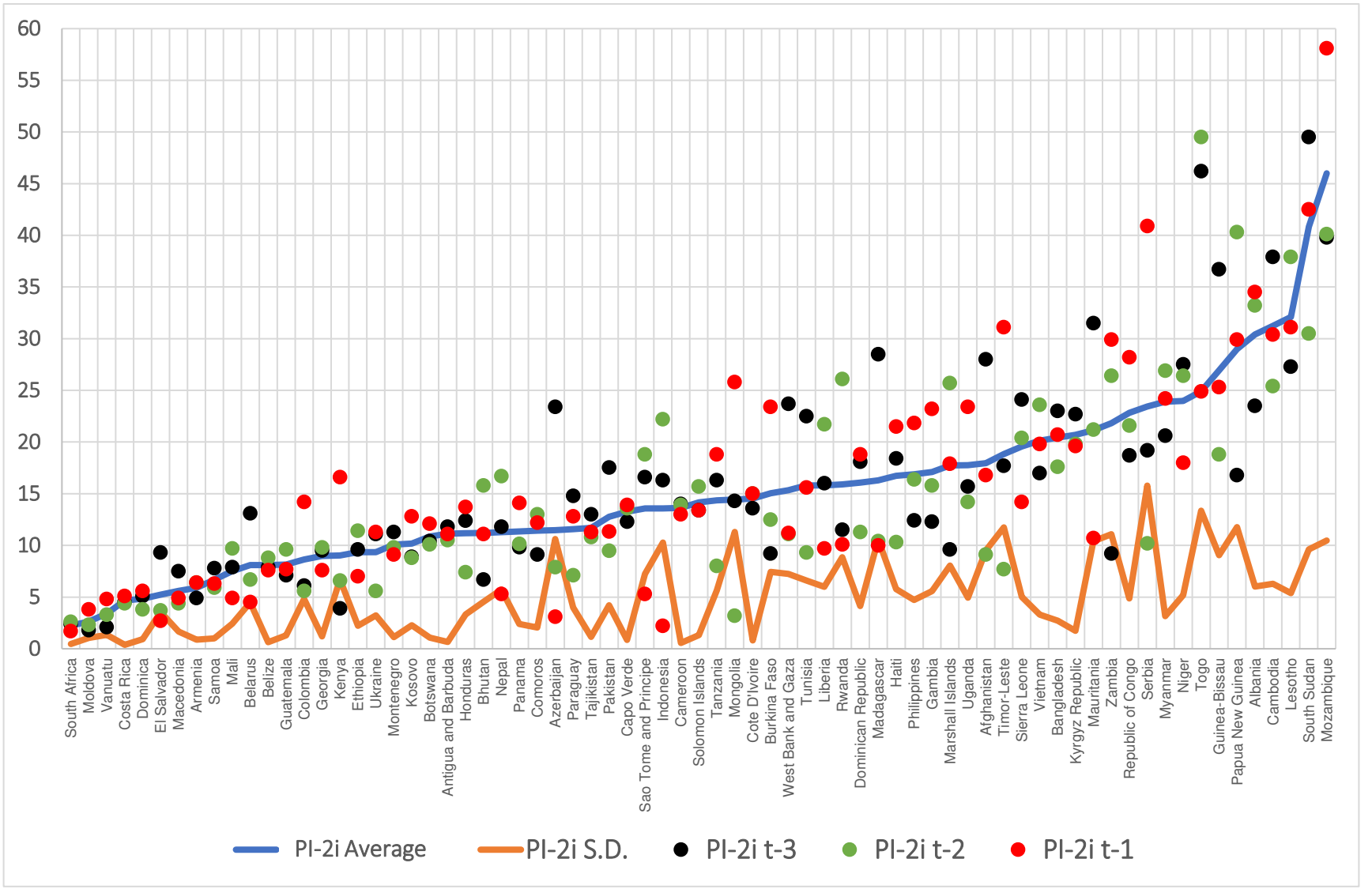
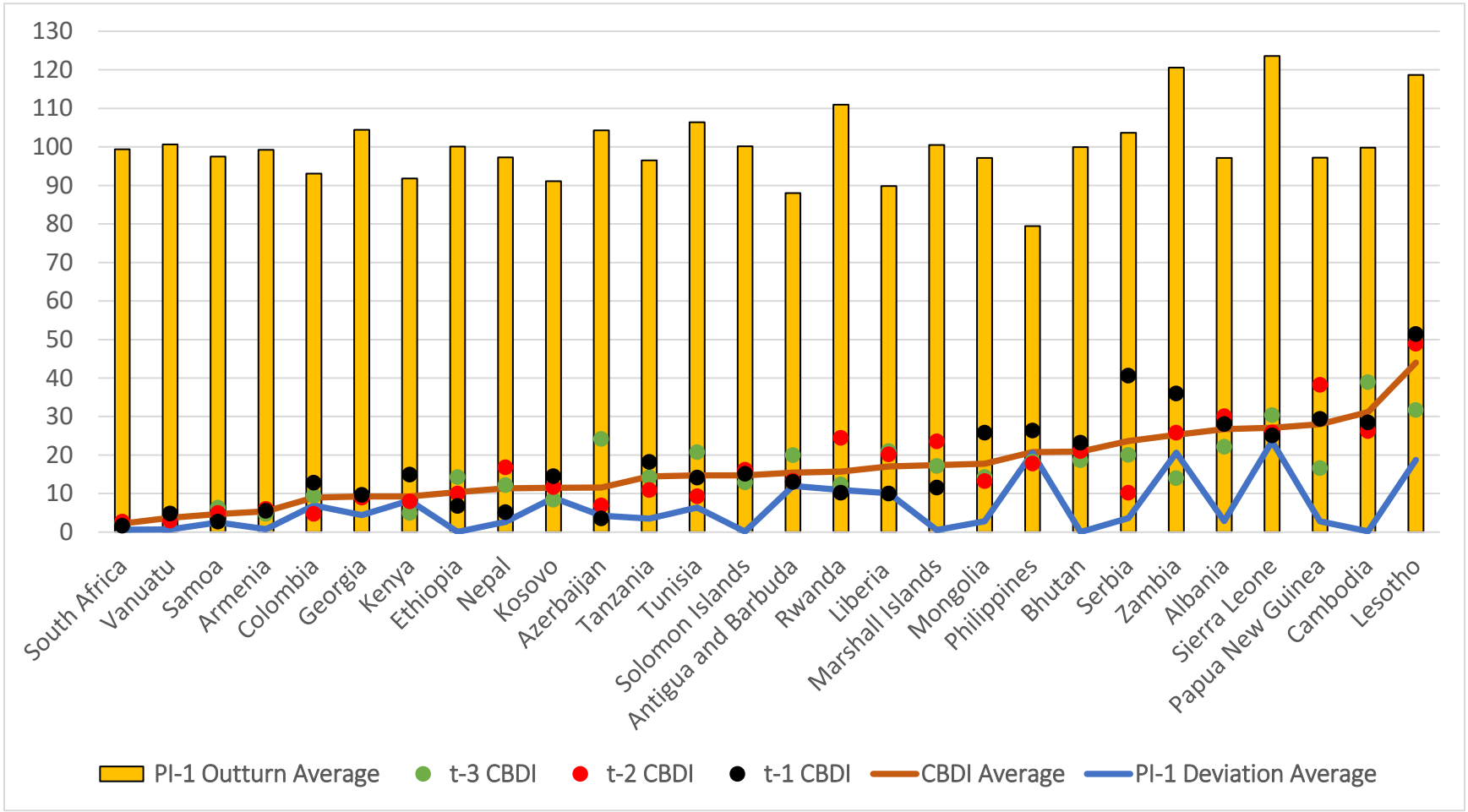


Figure 4. COMPOSITIONAL BUDGET DEVIATION INDEX (CBDI) VS. PI-1 BUDGET DEVIATION



DEVIATIONS IN AGRICULTURE, EDUCATION, HEALTH, AND PUBLIC WORKS MINISTRIES

This section seeks to understand whether key social sector ministries are proportionally or disproportionally affected by the execution of the total budget.

Figure 5 illustrates the average outturn for the agriculture, education, health, and public works ministries against the average outturn for the aggregate budget (PI-1 in black). The outturns for these sectors do not follow any clear overall pattern and seem to be country specific. There are, however, significant differences between the sector outturns and the aggregate budget outturn, especially in agriculture and public works.

This finding is further explored in Figure 6 which shows the standard deviation for the key sectors. Here, the gaps are more divergent as agriculture and public works, and to a lesser extent health, have standard deviations that are significantly above the standard deviation for the overall budget. Education, which is predominantly comprised of teaching staff, is the lone exception.

To understand the relationships between the aggregate budget outturn and the key sectors, the correlation matrix for the full dataset (three observations each, treated independently, rather than the country average) is shown in Table 8. Public works, comprising a significant portion of the capital investment budget, is the most highly correlated with the aggregate outturn, while the others are less so. Thus, the outturns in the agriculture, education, and health sectors appear to be only weakly related to the aggregate budget outturn.

TABLE 8. CORRELATIONS MATRIX FOR KEY SECTORS AND THE AGGREGATE OUTTURN

	PI-1	Agriculture	Education	Health	Public Works
PI-1	1				
Agriculture	0.41	1			
Education	0.23	-0.02	1		
Health	0.15	0.20	0.07	1	
Public Works	0.65	0.29	0.19	0.18	1

EXPLANATIONS FOR DEVIATIONS IN PEFA REPORTS

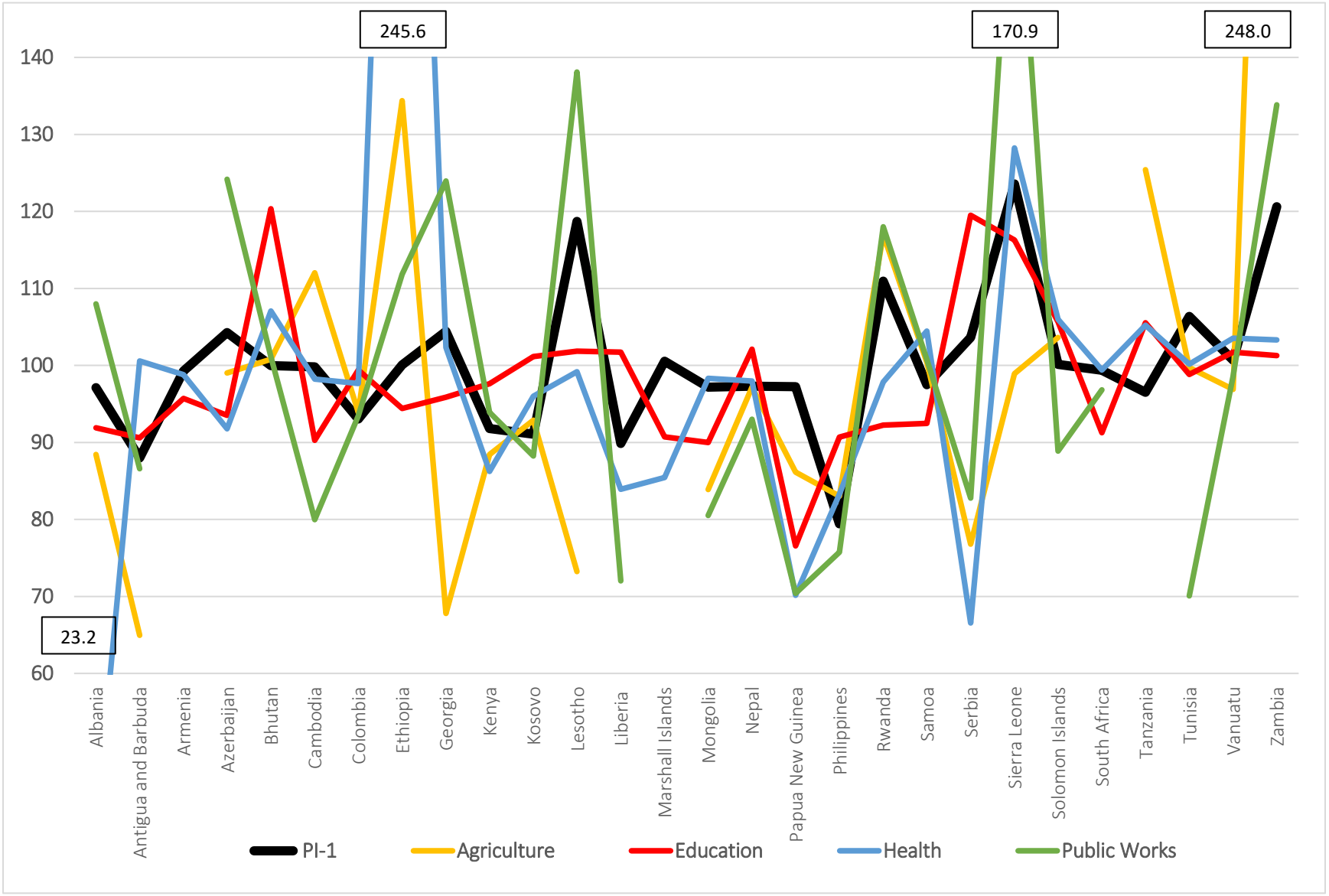
The PEFA reports vary significantly in the breadth of their explanations for these deviations. For some reports, there is a lengthy description justifying why budget execution may be under, as planned, or over for each of the

three years. At the same time, other PEFA reports only provide the data calculations and the scores, with little to no additional analysis.

Economic shocks were the most cited justification, followed by poor revenue forecasts, challenges in the budgeting process, and the use of supplementary budgets (see Table 9 and Annex 2 for an explanation of the deviation codes). On the revenue side, the most cited issue was challenges in revenue forecasting, followed by economic shocks, and the unreliability of donor funds.[13] "Not provided", the catch-all term used whenever there was not deemed to be sufficient justification, was the most common explanation for both PI-1 (22 PEFA reports) and PI-3 (28 PEFA reports)

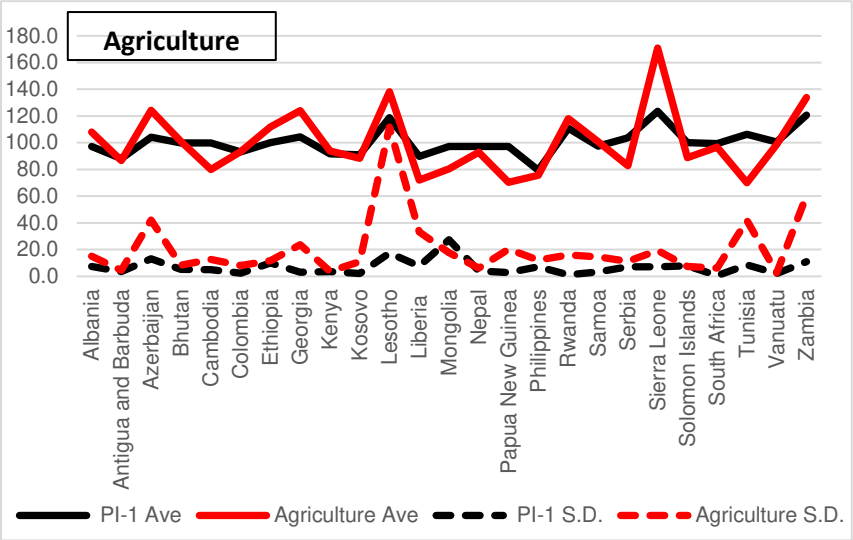
[13] Donor funding is including in the calculation of the revenue outturn in the PEFA framework. For a handful of countries in the sample, their poor rating on revenue mobilization was in large part due to donor funds being delayed or not forthcoming (for example Bhutan who scored A on PI-1 and D on PI-3).

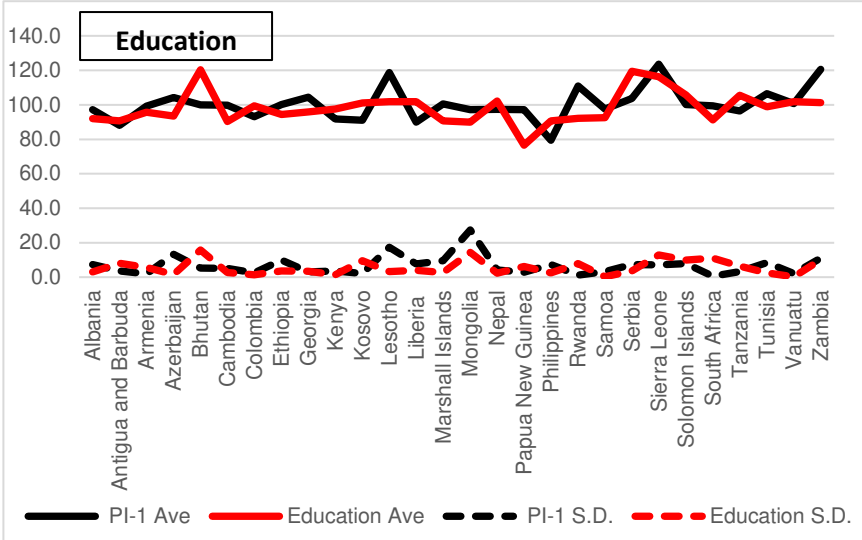
Figure 5. AVERAGE OUTTURN IN KEY SECTOR MINISTRIES

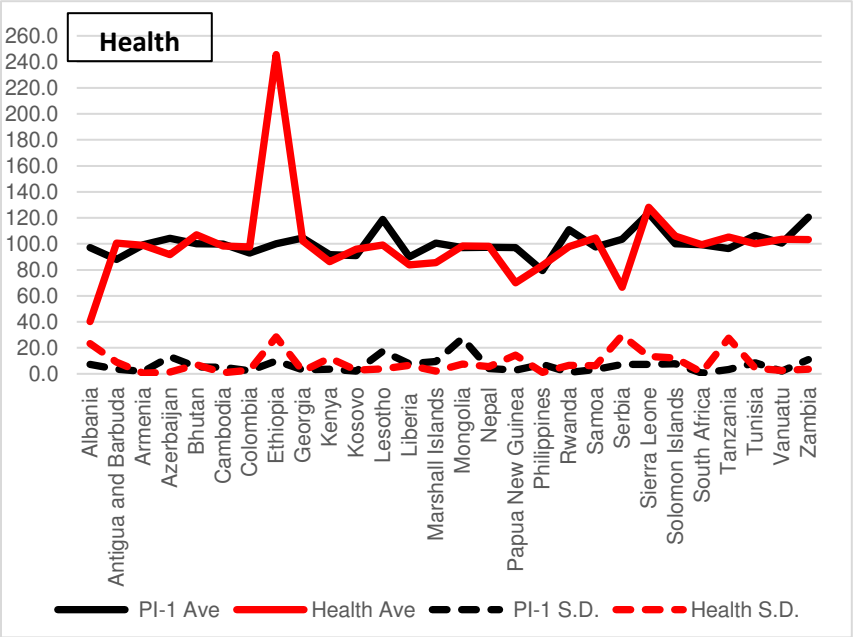


Breaks in the lines are due to missing observations; the numbers in the boxes are the values that were cut off by the axis range

Figure 6. AVERAGE OUTTURN AND STANDARD DEVIATION OF KEY SECTOR MINISTRIES







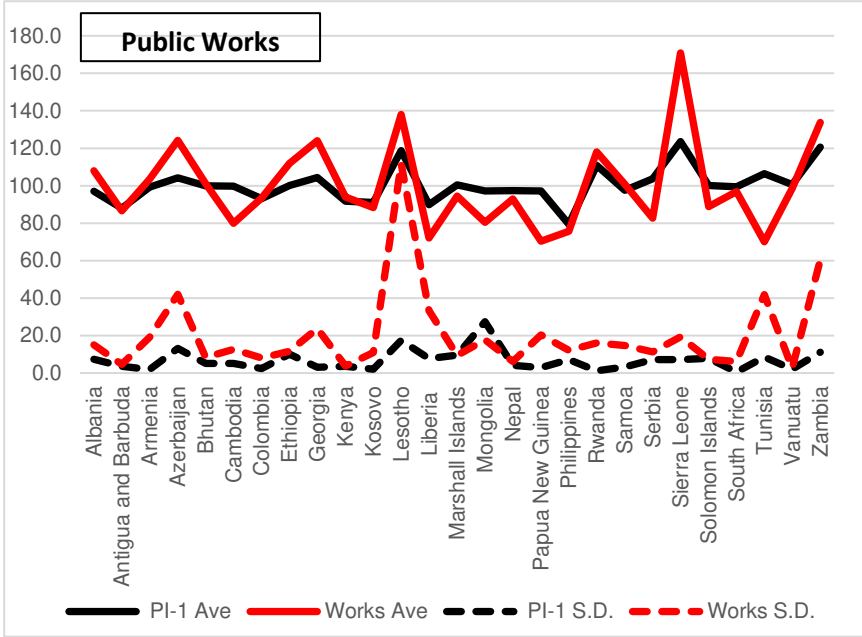


TABLE 9. EXPLANATIONS PROVIDED FOR DEVIATIONS

PI-1 Expenditure Deviations			**PI-3i Revenue Deviations**	
Deviation Code	*Count*		*Deviation Code*	*Count*
Not provided	22		Not provided	28
Economic	14		Revenue Forecasting	28
Revenue Forecasting	15		Economic	24
Supplementary Budget	10		Donors	5
Capital Budget	6		Revenue Administration	4
Donors	6		Exceptional Weather	2
Budget Institutions	6		Conflict	1
Expenditure Forecasting	4		Political Instability	1
Exceptional Weather	3		Revenue Policy	1
Political Instability	3			
Procurement	3			
Data reliability	2			
Conflict	2			
Borrowing	1			
Government Reorganization	1			
Refugee Crisis	1			
Health Emergency	1			
Total	***100***		***Total***	***95***

Note: Coding described in Annex 2

IS EXPENDITURE DEVIATION AFFECTED BY REVENUE DEVIATION?

From a theoretical perspective, the expenditure outturn could be highly dependent on the revenue outturn. For example, if revenue is short of expectations, will a government curtail expenditure or draw resources from other sources (say, borrowing) to meet expenditure targets? Also, if revenue exceeds expectations, will a government increase its expenditure or save these resources for future use? Several PEFA report write-ups noted the global financial crisis as the source of the economic shock that had an adverse impact on revenue collection, which in turn affected the ability to implement the budget as originally planned.[14] In other cases, where countries scored an "A" on PI-3, there were other factors that contributed to an inability to implement the budget. In Madagascar, for example, which scored D on PI-1 and A on PI-3, the PEFA reports points to the economic crisis that led to the

[14] The period of assessment in many PEFA reports occurred during the global financial crisis.

inability to spend the budget, but there is little explanation on PI-3 other than to say that they met their domestic revenue targets. This is evidence of incomplete explanations.

Figure 8 illustrates that countries score slightly worse on PI-3i than PI-1, but much better than PI-2i. In general, there is more variability around the 100 percent outturn level as fewer countries have an average outcome close to forecasted revenue. Also, there is a higher average standard deviation than PI-1 (8.5 vs. 7.6), which indicates less consistency from year to year.

To examine whether there is any relation between these revenue and expenditure outturn, a series of statistical analyses were performed. First, a graphical analysis illustrates the cumulative probability distribution of PI-1 and PI-3i for both the averages and for the individual periods of t-3, t-2, and t-1 (see Figure 7).

Figure 7. CUMULATIVE PROBABILITY DISTRIBUTION OF PI-1 AND PI-3I

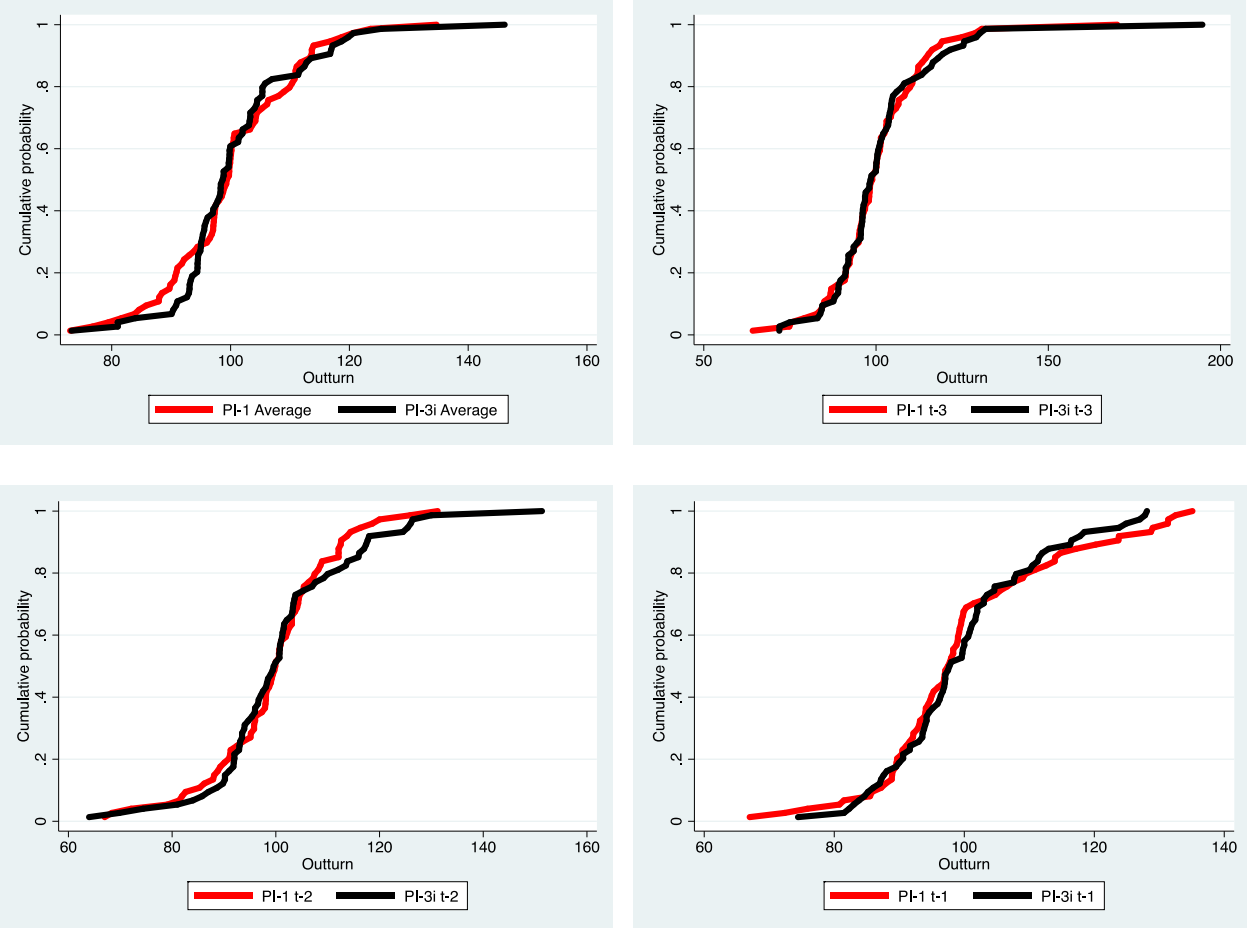
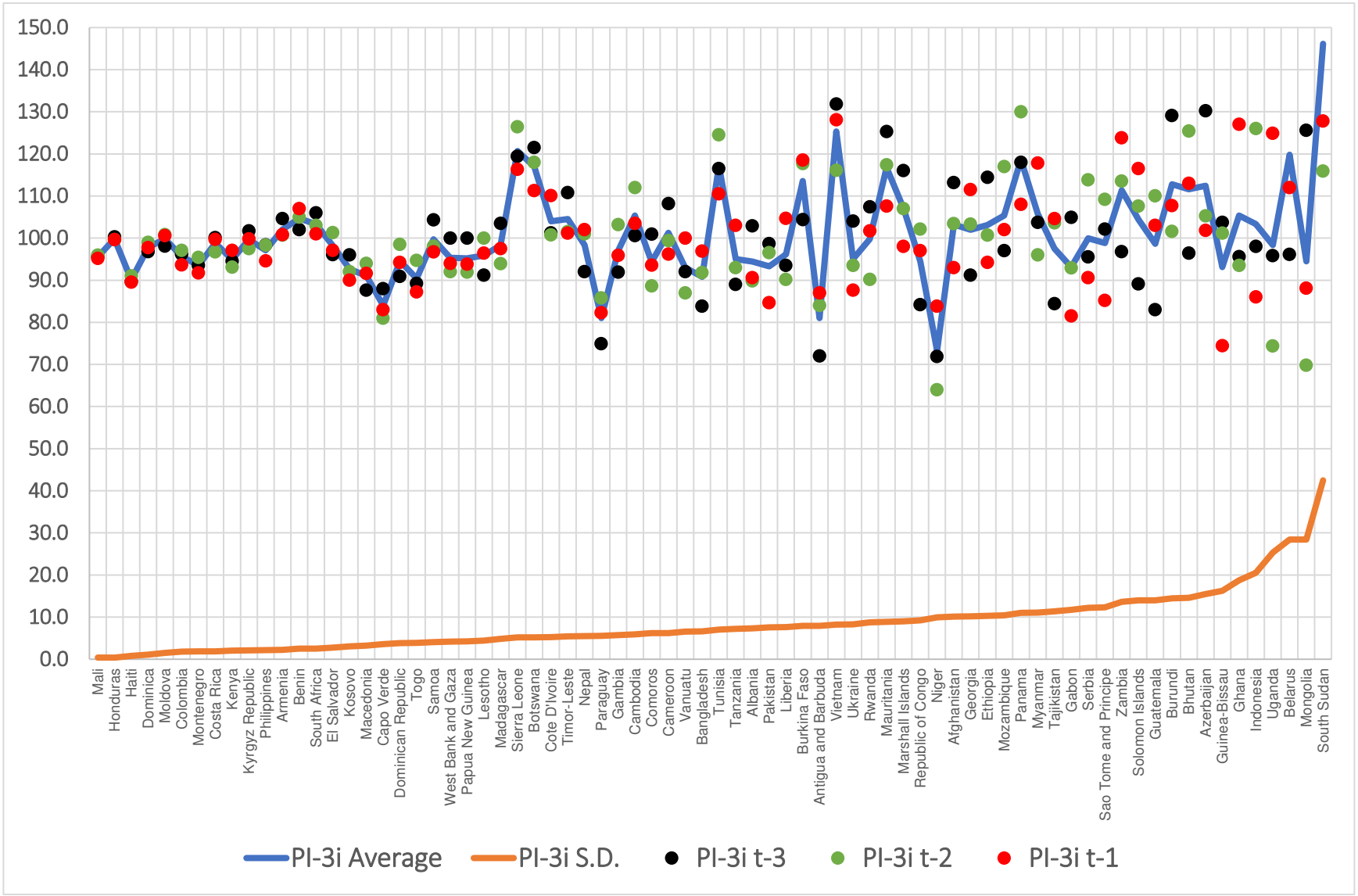


Figure 8. PI-3I REVENUE OUTTURN DATA POINTS, AVERAGE, AND STANDARD DEVIATION



Note: South Sudan had its t-3 observation at 194.7, which extends beyond the chart.

The cumulative probability distributions organize data from lowest to highest and plot the frequency of observations. Consequently, the relationship between expenditure and revenue within a country is lost since revenue and expenditure are both sorted separately. Nevertheless, this high-level analysis indicates that these two variables do not operate completely in sync and that there is a significant separation at times.

Second, a comparative analysis was conducted of the difference between budget and revenue outturns in percentage points for each year of analysis.[15] This method was chosen to analyze the relationship between expenditure and revenue. For example, if revenue exceeded the target, was there a corresponding increase in expenditure or was this extra revenue used to pay down the public debt? Similarly, if revenue was less than the target, was expenditure reduced or was money borrowed to maintain planned expenditure targets? Therefore, if countries have dots close to the zero line, it suggests that expenditure and revenue outturns are closely aligned and in sync.

The findings in Figure 9 show that a substantial portion of countries have a consistent and significant divergence in the gap between their expenditure and revenue outturns. The observations were color-coded to differentiate between observations where revenue exceeded expenditure (net savings) versus observations where expenditure exceeded revenue (net borrowing). Nearly a quarter of observations (52 of 221) had expenditure and revenue outturns that were within 2.5 percentage points, but a third of the observations (70 out of 221) had a difference greater than 10 percentage points and there were 40 observations with a difference greater than 15 percentage points (see Table 10). Interestingly, the Capo Verde (-24.3, -32.7, -30.9)[16], the Dominican Republic (-24.3, -32.7, -30.9), the Gambia (-12.6, -11.2, -35.5), Haiti (-16.8, -21.1, -10.4), and Lesotho (-7.7, -25.9, -34.9) all had a consistent percentage point spread in favor of greater expenditure than revenue. Conversely, Botswana (17.2, 19.0, 14.2), Cote D'Ivoire (10.0, 18.2, 18.5), Gabon (19.9, 25.9, 14.5), the Philippines (11.6, 19.2, 22.2) had a consistent percentage point bias of revenue exceeding expenditure. The largest percentage point difference occurred in Indonesia's t-2 year, where the revenue outturn was 126.0 and the expenditure outturn was only 72.1.[17]

[15] Thus, there are 74 x 3 = 222 observations, minus 1 (since the Republic of Congo is missing the t-3 observation) for a total of 221 observations.

[16] The difference between revenue and expenditure in percentage points, with negatives where expenditure was more than revenue (deficit) and positives where revenue was greater than expenditure (surplus).

[17] The PEFA write-up noted the under-execution of the capital budget in that year and the challenges in forecasting the price of oil and gas from which Indonesia derives 25 percent of its revenues.

Figure 9. EXPENDITURE AND REVENUE RELATIONSHIP, PERCENTAGE POINT DIFFERENCE IN EACH YEAR

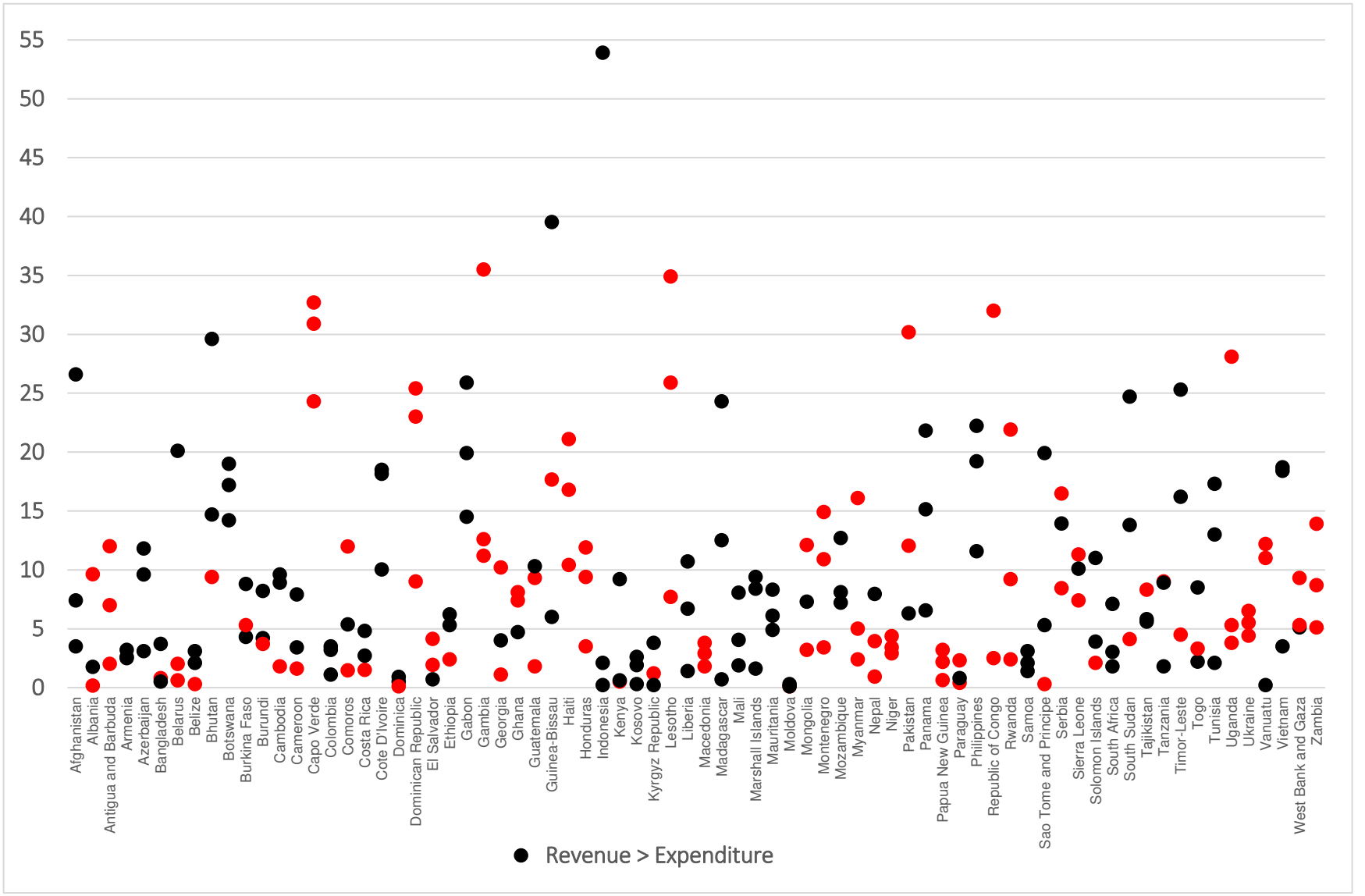


TABLE 10. DIFFERENCE BETWEEN EXPENDITURE AND REVENUE OUTTURNS

	0 to 2.5	2.5 to 5	5 to 10	10 to 15	> 15
Expenditure > Revenue	28	18	23	15	17
Revenue > Expenditure	24	22	30	15	23

Note: Column titles are the range of percentage point differences and the numbers in the cells represent the number of countries.

Third, a further statistical analysis investigates the correlations between PI-1 and PI-3i (see Table 11). For the period t-3, a high correlation of 0.74 was recorded though this number goes down slightly at period t-2 to 0.44 before rising slightly in period t-1 to 0.55. There is a similar high correlation of 0.58 between the PI-1 and PI-3i average outturns, and within PI-1 and PI-3i, suggesting a clear trend effect.[18] However, without a consistent set of years in the sample, it is a challenge to make any assumption regarding possible shocks or other explanation factors for why the correlations are slightly weaker between PI-1 and PI-3i in periods t-2 and t-1. So while the correlations point to a relationship between both indicators, more research will be needed to delve deeper into each country's specific circumstance and determine why the correlations are not consistent from year to year.

TABLE 11. BIVARIATE CORRELATIONS BETWEEN PI-1 AND PI-3I AT DIFFERENT TIME PERIODS

	pi1t3	pi3it3	pi1t2	pi3it2	pi1t1	pi3it1
pi1t3	1					
pi3it3	0.74	1				
pi1t2	0.50	0.27	1			
pi3it2	0.28	0.34	0.44	1		
pi1t1	0.29	0.01	0.74	0.19	1	
pi3it1	0.44	0.40	0.48	0.44	0.55	1

Collectively, the analysis presents a mixed picture for whether expenditure is dependent upon revenue. While the bivariate correlations indicate a high degree of correlation for t-3, the percentage point difference in Figure 13 shows that this holds true for a certain subset of countries, but for others there is a gap in revenue and expenditure outcomes. A longer period of analysis in each country that extends beyond 3 years will likely be needed to answer this question of codependence more satisfactorily.

[18] The average PI-2i score is only weakly correlated to the PI-1 average (0.09) and the PI-3i average (0.15).

IS THE DISTRIBUTIONAL IMPACT OF LACK OF CREDIBILITY DISCERNIBLE FROM THE PEFA REPORTS?

No. The PEFA reports do not discuss the distributional impact of deviations in expenditure from original planned appropriations. The Annexes that provide the compositional outturn data in each of the PEFA reports do provide data on whether the allocation of expenditure to key social sector ministries deviated from their planned budgets at the aggregate level. But allocation should not be confused with distribution. It is impossible to tell from these ministry level numbers whether one segment of society was affected more than another. At least one more level of disaggregation, and possibly more, would be required to start probing distributional impact from deviations in planned expenditure.

MAIN CONCLUSIONS

Budget credibility at first glance seems like a straightforward issue: was the budget executed as planned and were revenues collected as forecasted? As this paper has shown, however, there are multiple layers to budget credibility, including consistency, compositional, and the relationship between expenditure and revenue.

The following key messages derive from the analysis:

- Budget credibility is a multi-dimensional concept that requires a multi-dimensional analysis beyond a simple yes or no answer to the question “Does country X have a credible budget?”

- Budget credibility is first an accuracy issue (whether targets are met), second a consistency issue (where targets are met year to year), and third a compositional issue (whether this accuracy and consistency is maintained at sub-aggregate levels of expenditure and revenue).

- Expenditure was executed slightly better on average than revenue.

- South Africa was the most accurate and consistent of the countries in the sample.

- Very few countries accurately and consistently executed their budget as planned, and there were many cases of significant deviations from year to year, suggesting the presence of institutional and capacity challenges.

- A small segment of countries consistently overspent or underspent during the three-year period of analysis, suggesting inaccuracies in forecasting methods or in response to certain incentives.

- The aggregate budget tends to be more credible than expenditure composition at the administrative or functional level.

- The budget outturns in the agriculture, education, and health sectors show little correlation to the aggregate budget outturn, whereas the public works sector was the opposite and highly correlated to the aggregate budget outturn.

- The descriptive analysis in the PEFA report write ups could benefit from a more standardized approach to explanation of deviations.

- Deviations in expenditure are somewhat correlated with deviations in revenue, though this correlation is not consistent across the period of assessment and does not exist for many countries.

- Additional research is needed to further understand budget credibility.

SUGGESTED FUTURE RESEARCH

While the PEFA reports provide some initial evidence on the main issues of budget credibility, the assessed time period and short nature of the write-ups did not allow for a thorough analysis. For this reason, further research is suggested in a select group of country case studies.

The selection of case studies should be driven by a consideration of high and low performers across the range of budget credibility measures, and have regional and income level representation. For example:

- ***Top Performers***

Country	PI-1	PI-1 Ave	PI-1 S.D.	PI-2	PI-2i Ave	PI-2i S.D.	PI-3i	PI-3i Ave	PI-3i S.D.
South Africa	A	99.4	0.6	A	2.2	0.5	A	103.3	2.5
Moldova	A	99.7	1.3	A	2.6	1.0	A	99.8	1.5
Costa Rica	A	96.8	1.5	A	4.7	0.4	A	98.8	1.9
El Salvador	A	99.9	0.8	A	5.2	3.6	A	98.1	2.8

- ***Lowest Performers***

Country	PI-1	PI-1 Ave	PI-1 S.D.	PI-2	PI-2i Ave	PI-2i S.D.	PI-3i	PI-3i Ave	PI-3i S.D.
Sierra Leone	D	123.6	7.2	D	19.6	5.0	D	120.7	5.2
Niger	D	76.8	9.6	D	24.0	5.2	D	73.2	10.0
Mongolia	D	97.2	27.4	D	14.4	11.3	D	94.5	28.4

- ***Consistent Over-Spenders***

Country	PI-1	PI-1 Ave	PI-1 S.D.	PI-2	PI-2i Ave	PI-2i S.D.	PI-3i	PI-3i Ave	PI-3i S.D.
Capo Verde	C	113.3	0.9	C	13.3	0.9	D	84.0	3.6
Rwanda	C	110.9	1.2	C	15.9	8.9	B	99.8	8.8
Vietnam	C	111.8	2.1	D	20.1	3.3	D	125.3	8.2

- ***Consistent Under-Spenders***

Country	PI-1	PI-1 Ave	PI-1 S.D.	PI-2	PI-2i Ave	PI-2i S.D.	PI-3i	PI-3i Ave	PI-3i S.D.
Kosovo	B	91.1	2.0	B	10.2	2.3	C	92.7	3.1
Togo	C	87.9	2.3	D	24.9	13.4	D	90.4	3.9
Colombia	B	93.0	2.4	B	8.6	4.8	B	95.6	1.6

- ***Revenue Outturn Significantly Greater than Expenditure Outturn***

Country	PI-1	PI-1 Ave	PI-1 S.D.	PI-2	PI-2i Ave	PI-2i S.D.	PI-3i	PI-3i Ave	PI-3i S.D.
Botswana	A	100.1	3.7	C	10.9	1.1	D	116.9	5.2
Cote D'Ivoire	C	88.5	5.1	C	14.5	0.8	B	104.0	5.3
Philippines	D	79.4	7.2	D	16.9	4.7	A	97.1	2.2

- ***Expenditure Outturn Significantly Greater than Revenue Outturn***

Country	PI-1	PI-1 Ave	PI-1 S.D.	PI-2	PI-2i Ave	PI-2i S.D.	PI-3i	PI-3i Ave	PI-3i S.D.
Capo Verde	C	113.3	0.9	C	13.3	0.9	D	84.0	3.6
Dominican Republic	D	113.7	5.4	D	16.1	4.1	B	94.5	3.8
Haiti	B	106.2	6.1	D	16.7	5.8	D	90.1	0.8
Gambia	C	116.8	13.6	D	17.1	5.6	B	97.0	5.7
Lesotho	D	118.7	17.4	D	32.1	5.4	B	95.9	4.4

- ***Top Performers in Expenditure Composition Credibility***

Country	CBDI	Agriculture		Education		Health		Public Works		4 Sectors	
	AVE	*AVE*	*SD*	*AVE*	*SD*	*AVE*	*SD*	*AVE*	*SD*	*AVE*	*SD*
South Africa	2.2	96.8	6.2	91.2	11.0	99.4	1.2	96.8	6.2	96.1	6.1
Vanuatu	4.6	98.6	2.6	101.7	0.2	103.5	2.5	98.6	2.6	100.6	2.0
Colombia	9.0	93.4	8.1	99.4	1.3	97.7	2.9	93.4	8.1	96.0	5.1
Kenya	9.3	93.9	3.8	97.6	1.3	86.2	12.1	93.9	3.8	92.9	5.2

Note: The 4 Sectors column represents the average budget outturn and standard deviation for agriculture, education, health, and public works.

- ***Lowest Performers in Expenditure Composition Credibility***

Country	CBDI	Agriculture		Education		Health		Public Works		4 Sectors	
	AVE	*AVE*	*SD*	*AVE*	*SD*	*AVE*	*SD*	*AVE*	*SD*	*AVE*	*SD*
Sierra Leone	27.1	170.9	19.3	116.3	12.9	128.2	13.5	170.9	19.3	146.6	16.2
PNG	28.0	70.4	20.4	76.6	6.1	70.2	14.5	70.4	20.4	71.9	15.4
Cambodia	31.2	80.0	12.7	90.3	2.8	98.2	0.9	80.0	12.7	87.1	7.3
Lesotho	44.0	138.1	111.7	101.9	3.3	99.2	3.7	138.1	111.7	119.3	57.6

Note: The 4 Sectors column represents the average budget outturn and standard deviation for agriculture, education, health, and public works.

Future analysis could build upon the information in the PEFA reports by considering:

- ***Longer time frame:*** A longer period of study (more than 3 years) would allow for a richer understanding of the issues and reduce the possibility of shocks skewing the data. Many countries in the sample were affected one way or another by the global financial crisis.

- ***Revenue forecasting methods:*** With a few exceptions, very few of the PEFA reports go into the details about how the revenue forecasts are undertaken. Are forecasts linked to prior year forecasts or prior year actual collections? Are forecasts undertaken for individual taxes or revenue categories?

- ***Decisions around budget execution:*** What happens where there is substantially more revenue than budgeted but no more spending? And vice versa?

- ***Impact on other variables*****:** It would also be interesting to understand the impact on (i) the budget deficit, (ii) arrears, (iii) the quality of financial statements, and (iv) whether budget statements are audited and publicly available to get a sense about the reliability of reported expenditure. For example, if a country

scores well on the budget outturn, but it's financial statement was not audited, would it be viewed in the same way as a country whose financial statement was audited?

- ***Understanding incentives:*** Are there incentives to under or over forecast revenue? And does this occur in the countries that consistently over-executed and under-executed their budget?

- ***More detailed analysis of compositional budget deviations at the country level:*** For example, are the approved budgets for certain ministries routinely under- or over- executed? It would also be helpful to analyze the reasoning behind changes in the compositional execution rate year to year and whether distributional implications could be understood.

ANNEX 1. DIFFERENCE BETWEEN 2011 AND 2016 PEFA SCORING METHODOLOGIES

	2011 Methodology		2016 Methodology	
PI-1 Aggregate expenditure outturn				
Scoring	A	Aggregate expenditure outturn was between 95% and 105% of the approved aggregate budgeted expenditure in at least two of the last three years.	A	In no more than one out of the last three years has the actual expenditure deviated from budgeted expenditure by an amount equivalent to more than 5% of budgeted expenditure.
	B	Aggregate expenditure outturn was between 90% and 110% of the approved aggregate budgeted expenditure in at least two of the last three years.	B	In no more than one out of the last three years has the actual expenditure deviated from budgeted expenditure by an amount equivalent to more than 10 % of budgeted expenditure.
	C	Aggregate expenditure outturn was between 85% and 115% of the approved aggregate budgeted expenditure in at least two of the last three years.	C	In no more than one of the last three years has the actual expenditure deviated from budgeted expenditure by more than an amount equivalent to 15% of budgeted expenditure.
	D	Performance is less than required for a C score.	D	In two or all of the last three years did the actual expenditure deviate from budgeted expenditure by an amount equivalent to more than 15% of budgeted expenditure.
Coverage	Budgetary central government		Budgetary central government	
PI-2i Expenditure composition outturn				
Scoring	A	Variance in expenditure composition exceeded 5 % in no more than one of the last three years.	A	Variance in expenditure composition by program, administrative or functional classification was less than 5% in at least two of the last three years.
	B	Variance in expenditure composition exceeded 10 % in no more than one of the last three years.	B	Variance in expenditure composition by program, administrative or functional classification was less than 10% in at least two of the last three years.
	C	Variance in expenditure composition exceeded 15 % in no more than one of the last three years.	C	Variance in expenditure composition by program, administrative or functional classification was less than 15% in at least two of the last three years.
	D	Variance in expenditure composition exceeded 15 % in at least two of the last three years.	D	Performance is less than required for a C score.
Coverage	Budgetary central government		Budgetary central government	
PI-3i Revenue outturn				
Scoring	A	Actual domestic revenue was between 97% and 106% of budgeted domestic revenue in at least two of the last three years	A	Actual revenue was between 94% and 112% of budgeted revenue in at least two of the last three years.
	B	Actual domestic revenue was between 94% and 112% of budgeted domestic revenue in at least two of the last three years	B	Actual revenue was between 94% and 112% of budgeted revenue in at least two of the last three years.
	C	Actual domestic revenue was between 92% and 116% of budgeted domestic revenue in at least two of the last three years	C	Actual revenue was between 92% and 116% of budgeted revenue in at least two of the last three years.

	D	Actual domestic revenue was below 92% or above 116% of budgeted domestic revenue in two or all of the last three years	D	Performance is less than required for a C score.
Coverage	Budgetary central government		Budgetary central government	

ANNEX 2. DEVIATION CODES

Economic: Macroeconomic shock to growth, inflation, etc. that resulted in less economic activity than anticipated

Revenue Forecasting: Whether optimistic or conservative revenue projections

Expenditure Forecasting: Whether optimistic or conservative expenditure projections

Budget Institutions: Any reason stemming from issues with the budgeting process, including fiscal rules, virements, commitment control, planning, and arrears.

Procurement: Challenges or delays in the procurement process

Capital Budget: Under-execution of the capital or development budget

Political Instability: All matters related to a political crisis

Conflict: When conflict was provided as justification for the deviation

Exceptional Weather: Floods, earthquakes, etc.

Supplementary Budget: Provided as a justification to explain why the expenditure outturn differed from original estimates

Donors: Cases where revenue forecasts were not met due to donor funds being less than expected or not forthcoming

Government Reorganization: Was used in one case to justify the deviation in expenditure and did not fit another category

Health Emergency: Relates to the outbreak of Ebola in Liberia

Data Reliability: Relates to expressed concerns about the quality of budget data in Lesotho and Tajikistan

Borrowing: Relates to unplanned borrowing from the domestic banking system in the West Bank and Gaza

Refugee Crisis: Relates to the influx of refugees into Burkina Faso stemming from the political crises in Guinea-Bissau and Mali

Revenue Administration: Explicit mention of issues concerning or reforms to the revenue administration agency

Revenue Policy: Explicit mention of changes in revenue policy

Not provided: Catch-all term for cases where there was not a justification to explain the deviation in expenditure or revenue, or where it was judged that there was not a sufficient explanation within the write-up provided (a strict scoring system was used whereby there had to be more than an explanation of whether one category of expenditure, for example, moved up or down, but rather the explanation of the root causes behind the movement)

ANNEX 3. TERMS OF REFERENCE

The consultant will analyze all the Public Expenditure and Financial Accountability (PEFA) country reports published since 2012 and focus on data and explanations contained in these reports on indicators related to budget credibility/deviations (PI 1-3 in the PEFA 2011 and 2016 frameworks). Only one report (the latest report) per country should be analyzed.

As part of the analysis, the consultant will create a database on budget credibility findings in PEFA reports. The database should be populated with the scores received by countries on each of the relevant PEFA indicators. Based on the scores on these indicators, the consultant should propose a country grouping approach based on the seriousness of the budget credibility problem. For example, countries could be categorized as those in which budget credibility is "not a big problem" (countries that receive A or B scores in two of the three PEFA indicators and those in which budget credibility is "a big problem" (countries that receive C or D scores in two of the three PEFA indicators), or other similar groupings. The database should be created in a manner that enables the consultant to distinguish countries in which budget deviations are caused due to overspending and underspending, and the extent of the over/underspending – e.g. as % of total government spending. This information will be available in the narrative sections of the PEFA country reports. The database should also include a column explaining why the deviations occurred (based on the information contained in the PEFA country reports).

The consultant will begin by the study process by developing a short methodological note about how the data will be collected and analyzed. This methodological note will be reviewed by IBP and finalized by the consultant based on IBP's comments. The final methodology will be followed by the consultant in developing the full report described below.

The consultant will write a 10-15 page report with a 2-page executive summary analyzing the database. The report should:

- Provide an overview of the extent of the budget credibility problem that exists globally focusing on the different categories of budget deviations that are assessed under the PEFA framework;

- Discuss the extent of budget deviations due to underspending and overspending, patterns on deviations emerging at the global or regional levels along with the causes for the deviations as described in the PEFA country reports.

- List the names of countries in which budget credibility is a problem due to overspending and underspending, based on the severity of the problem (countries scoring worst should be listed first) and on the type of reasons given for the deviations.

- Analyze whether the explanations for deviance in PEFA narratives are adequate

- Analyze whether PEFA's overall assessment of credibility is comprehensive enough to allow IBP to understand distributional implications.