



Making, Keeping and Breaking Promises: a global review of government budget credibility

Jason Lakin, Ph.D.
Strathmore University: Public Lecture
May 16, 2019

How should we think about credibility?

Budget credibility refers to the difference between a government's approved budget and its actual performance: high credibility means that there are few differences between the budget and actual revenue and spending, while low credibility means that there are many deviations from the budget.

We can think of credibility from a purely PFM perspective, as a technical problem. Are we able to properly project our revenue and expenditure? Are these projections based on sound principles and past experience?

We can also think of credibility as a management problem: are we able to control our revenue collectors and line ministries/managers during the budget year? Do those who are authorized to commit funds actually do so in line with the budget and in a timely fashion? Do they have the skills and incentives to do so? Do those who are supposed to collect revenue do so?

We can also think of credibility as a political problem: can we maintain commitments to live within our means during the year in spite of pressures to drive up the deficit? Also, can we sustain political commitments to prioritize spending on certain sectors or projects against pressure from civil servants, interest groups, contractors or even citizens to deviate from our commitments during budget execution?

Finally, we can think of credibility as part of the democratic pact that we have in a society in which there is some kind of social contract between taxpayers and the state, or more broadly between the governed and the governors. In this case, the budget represents a kind of promise to citizens and the question is whether the state makes realistic promises and keeps them. As it will inevitably fail to keep some of them, this idea gives us a related idea that there must be some form of dialogue around changes to the pact that occur during and after budget execution- some requirement to publicly justify or explain the changes.

What difference does credibility make?

If a budget is not realistic and it is underspent, so what? If we all know that the budget is "padded," we should just discount the numbers at the beginning of the year, and we will not be surprised at the end of the year. So what? Let us start with a fact: from our analysis of 35 countries around the world (for which we have 2-9 years of data each=232 country-years of data), mainly low and middle income countries, the average underspending across these budgets overall is 9 percent in a given year. This is an aggregate, and we can imagine a scenario where the budget is systematically underspent every year by that same 9 percent (that is just a hypothetical, but it would be consistent with our data).

So what? Presumably everyone knows that this is happening, and they simply discount the budget by 9 percent at the start of the year.

On some level, this is right. If the patterns are consistent and are, for example, mainly about making the budget appear to be more capital than recurrent, and yet we know the truth, we need not worry too much. Of course, it is the case that by promoting such fictions, we avoid confronting the problems we have: at least some people will look at the budget and think that the allocations look right, even though they are not realistic. But even those who know better may indulge in other fictions, such as the idea, prevalent in Kenya, that the recurrent budget is intrinsically wasteful and if only we spent more on capital, we would not have the problems we have.

This is not right, but to see this, we must confront the recurrent budget more carefully and ask why we don't seem able to reduce it. We need to look at what is worthy and what is wasteful in the recurrent budget. But we may not have this discussion if we cannot: when the budget is presented, it will always seem to have solved the problem. We may know the budget is not realistic, but we also know that we cannot see the real problems in it.

The bigger issue, it seems to me, is that the patterns of poor credibility are rarely consistent, and they create manifold opportunities to undo what is done in budget formulation during budget execution. This is an altogether different matter and of far more significance.

Here we have in mind the decision to reallocate funds from agreed priorities to others, such that a 9 percent decline in the overall budget is not distributed equally, but in ways that favor particular areas and not others. We call these compositional credibility problems, that is to say, the difference in the relative share of the budget that goes to different areas in the approved versus the executed budget.

Of course, we have already spoken about one way in which this happens at the level of the economic classification of the budget, which is a shift from capital to wages. And this is indeed born out by the data we have at global level. For our 35 countries, capital expenditure is below budget by 9 percentage points on average (net of overall changes in the budget), while recurrent expenditure is overspent by 4 percent (net) and wages/compensation by nearly double that. Another way of saying this is that (if we remove a small number of outliers defined using standard box-plot interquartile range thresholds), the share of the budget going to recurrent increases by about 3 percentage points on average during budget execution, while capital falls by the same amount.

Again, it might be that these kinds of changes happen every year, and that they are entirely predictable, but this becomes less likely as we disaggregate the budget further. It becomes harder to be certain whether and by how much a particular part of the budget really will change during the year, and this makes it difficult to debate tradeoffs.

Compositional credibility is not just about economic classes, but about functions. Here again, we have some global data we can draw on that tells us something about this problem. From our 35-country sample, if we look at spending by functions, we know that:

- Economic affairs, environmental protection and housing & community amenities are underspent on average...
- ...while defense (for BOOST countries with fully public data available), public order & safety, social protection, etc. tend to be overspent. Health is also underspent on average, although not by much.

These relationships are somewhat attenuated in cases where the budget is overspent in the aggregate (by more than 5%). Though relatively uncommon in our sample, in these cases, there tends to be more capital spending and more spending on economic affairs, a capital heavy function.

Underspending on economic affairs particularly impacts agriculture, which tends to be the most underspent of three major sub-sectors (ag, transport, fuel/energy) available. In the smaller set of countries for which we have available data (13), we can see that within agriculture, irrigation is a major driver of underspending. In our larger sample, agriculture is underspent by about 2.5 percentage points (net of overall economic affairs expenditure and removing outliers). The underspend in agriculture in our smaller sample where irrigation is also visible is larger at nearly -6 percentage points (net, removing outliers). But irrigation in this sample is underspent on average by around 17 percentage points (net, removing outliers).

Kenya happens to be one of these countries. Between 2013/14 and 2016/17, irrigation underspending in Kenya amounted to roughly 38 percent of the budget (raw). Irrigation underspending is mainly concentrated in capital spending, and less than half of the total budget for capital assets was spent in this period. Large capital projects have been consistently underspent in recent years, including the Galana Kulalu agricultural modernization project and the Mwea irrigation (Thiba Dam) project. In 2016/17, these two flagship projects were each underspent by 45.4 and 61.4 percent and together accounted for 55 percent of total irrigation underspending.

The data we have is typically a bit shaky and limited, but it seems clear that this underspending is connected with inability to meet nonfinancial targets for the sub-sector. For example, in 2016/17, the Galana project only reached half of its target for acreage covered in the model farm, and Mwea reached less than half of its irrigated acreage coverage target.

Another area that is systematically underspent globally, in the reduced sample for which we have it, is vocational education. In 18 countries for which vocational education is available as a subsector, vocational education is underspent the most, by about 10% of the budget on average (raw, excluding outliers). Relative to the overall education budget, the subsector is underspent by about 6.5 percent (net, excluding outliers). We have not looked into this in further detail. As it happens, it is not an area of particularly low credibility in Kenya.

From the perspective of economic growth or development, we may well be shooting ourselves in the foot every time we fail to deliver on our promises. But we are also, from the perspective of the democratic pact, undermining confidence in government and the idea that there exists a public interest that is not simply the plaything of self-interested government officials.

What else do we know about budget credibility?

Let's talk about the Nigeria case, one of our first country case studies under the credibility project. Nigeria has a long history of mismanagement, rooted in a culture and economy that is organized around well-known pathologies of extractive industries. Nonetheless, we felt that there was more to say about Nigeria's credibility challenges, and at least some data to start looking at it.

Our overall story is that yes, as one would expect, some part of the credibility problem in Nigeria is caused by fluctuating oil prices, though in most cases, poor forecasting of oil revenue in Nigeria reflects mainly poor estimation of production rather than price: Nigeria is reasonably conservative in estimating oil prices each year. So, we start already with a recognition that even in this area where the problem may seem at first sight to be structural, there is more policy choice than is sometimes recognized.

But the reality is more damning still, because to the extent that Nigeria's credibility problem is tied to revenue, the revenue shortfall in most years is not in its majority caused by oil revenues. Yes, oil is important. But other revenue sources account for a majority of the revenue shortfall in most years between 2009 and 2016, including the spectacular failure to martial independent revenue collected by state agencies, which was by far the main cause of the revenue shortfall in 2016. Customs revenue is a consistent driver of poor revenue performance in Nigeria.

Yet the credibility problem in Nigeria is also about ambitious expenditure estimates, regardless of revenue figures, and these are fueled, in at least some years, by legislative amendments to the capital projects budget. The introduction of these projects without funding or feasibility studies tends to result in underspending. However, the role of the legislature is also not uniform. In some years, it has actually pushed for a more realistic budget.

Finally, while the bulk of the Nigeria problem is about poor budgeting at the front end, there is also evidence of underspending by ministries and agencies of the funds released to them. This points to an important distinction in the analysis of credibility, which is between problems in budget formulation and problems in budget execution. Unrealistic budgets are related more closely to formulation problems, while failure to spend available funds by line ministries is more related to execution. Of course, the line is not a red one: poor execution of available cash may still be about poor budget formulation, and high execution of available cash but low overall budget execution could be about ministries receiving funds in the last month of the year and being unable to spend them, which may be more about cash management than budget formulation. In any case, we at least know that when there is no cash, we are not speaking about an execution but a formulation problem.

We are in the early days of the Philippines case study, but it is interesting because it is a country where there has been a very substantial increase in budgeted resources for social sectors, but execution remains a problem. The data suggests that variation in the capacity of regions to spend may also exacerbate inequalities, and this is one neglected area of credibility studies. What is the relationship between credibility and equity? There are at least two hypotheses here:

1. Reallocation of funds during the budget year could lead to inequity.
2. Inability to execute budgets even if they are redistributive could lead to more inequality or a failure to reduce inequality.

The first hypothesis requires some detailed assessment of the kinds of changes that are made during the year. One way to look at this is to assess project expenditure that can be tagged to specific regions for changes between budget and execution. There is not a lot of data here to assess this, but it is theoretically possible. We are looking at this in Nigeria to see if we can track it.

Our second hypothesis looks more at the units that receive funds, which could be subnational units or deconcentrated national agencies. Our Philippines case study looks at variations in the ability of regional offices of the national Department of Health to spend their budgets, particularly their operational and capital budgets, and the implications that this may have for addressing inequality through social sector spending. We will have results in a few months.

We can look at this in Kenya too: are the counties with the worst budget credibility also the most marginalized? Data is available to look at this, though it requires some cleaning due to different ways of reporting. Interestingly, the data we have do not show any clear pattern of this type. We are still cleaning it, so I don't want to say anything definitive, but you can see the top third of counties in terms of development budget execution here, averaged for the period. Clearly some of our more marginalized counties actually perform well, though not all of them are here.

How governments should talk about credibility

Again, from the perspective of a democratic pact, governments have to share information about their deviations from budget and the reasons for these.

We investigated this matter in a study with 24 civil society organizations working in 23 countries at the end of 2018. We asked them to identify key programs of interest in health or education or other areas that were priorities. They looked at women's programs, capital budgets, agriculture and even revenue in addition to health and education.

What did we find? In fact, a majority of governments provide some kind of reasons for budget deviations in documents. But the standards for justifying or explaining themselves are exceedingly low. We start with some basic distinctions. We found two types of reasons to be common, and inadequate. The first is what we will call general. This is where a reason is given that is meant to encompass the overall deviation of the budget, but which does not (cannot) address micro-variation. For example, if we say that the budget was reduced by 10 percent because of a macroeconomic shock, that is fine. But it is only a specific reason if every ministry was also reduced by 10 percent. If, on the other hand, and as is usual, the overall decline leads to a shift in the relative share of the budget going to different areas (a compositional change), this is a "general" explanation that does not offer specific reasons for the changes we are interested in.

A second type of reason is what we may call generic. Here, the reason is meant to be specific to the area of the budget concerned, but it is too generic to offer a clear causal explanation. For example, we might say that the education budget is underspent because of poor planning. This might be true, of course, in some broad sense. But it is too generic: what does it mean and what exactly is the nature of the planning problem that directly causes the underspending problem? Another way to see this is that poor planning could explain both underspending and overspending. On its own, we cannot distinguish between these possibilities. An explanation that could explain very different outcomes is too generic to be useful.

Most of the reasons we found were either general or generic.

In our study, we tried to set some stricter criteria. For example, we tried to insist on a specific causal mechanism, which is a way of avoiding "generic" reasons. We also asked that reasons be able to explain variation over time or across similar areas. A typical generic reason explains nothing in a particular year, nor any variation over time: consider this table from Paraguay. What can explain these patterns of behavior-- increases and decreases over time, and differences in the patterns across regions? Certainly not "national bureaucracy", as a couple of subnational governments claimed.

Our criteria were useful, but also difficult to apply consistently to a wide array of problems that require nuanced policy understanding to diagnose. Nevertheless, the need for specific mechanisms and for reasons that are consistent with variation over time and across sub-sectors seems important, and we were able to discriminate among better and worse explanations using these ideas.

We are also looking at the problem from additional angles. For example, we are investigating the standards for governments to justify themselves more broadly, like those captured in the idea of "fair administrative action," which is something that Kenya has put in its laws. I will however give an example from a recent court case in the US regarding policy decisions found to be "arbitrary and capricious." These also give some standards that we can use. Note that there is an important conceptual difference here: in our original approach, we asked what would make a reason "adequate."

Administrative law starts from a negative proposition, asking not when a reason is adequate, but when it is inadequate, which means here “arbitrary” in some fundamental sense. This strategy has some advantages, as it does not require us to rule in reasons, but only to rule them out.

We draw on American case law here (see *Stewart v. Azar* 2018 on the Kentucky Medicaid waiver). Fair administrative action requires “reasoned decision-making.” Clearly, this means that there must be reasons of some kind involved and that these reasons are public or published.

What else? No important aspect of the problem can be entirely ignored in reasoning. Also important is **evidence** of consideration: as the court avers “[s]tating that a factor was considered ... is not a substitute for considering it.” There must be some evidence that factors were actually considered. That is to say, if a factor is dismissed, why was it dismissed? Nor can the fact that some rational basis for a decision can be construed after the fact sufficient; the reasons must have been actually constructed and provided at the time the decision was made. I interpret this as a transparency requirement; we are not to rely on guessing at matters of state. They must be revealed publicly and in good time if they are to count for anything.

Case law yields further insights into how the reasoning of the state may and may not be constructed to avoid being classified as arbitrary. We cannot enter into all of them here, but I will just highlight one idea from the same Medicaid case, which is that administrative decisions must be consistent with the government’s political and legal commitments as embodied in laws.

In the Medicaid case, the legislature passed a law whose objective was to increase access to health care. When a public official decided to adopt a policy that led to a reduction in access to health care services through Medicaid on the grounds that forcing people to work as a condition of receiving health services was good for population health, the courts rejected this reasoning because the official “cannot choose his own means to [the legislatures’] end.” This is analogous to deciding to shift the budget from health to defense during the budget year because defense will improve health. The point is not that this is necessarily false; the point is that if the legislature approved funds for health and funds for defense in the budget, changing those allocations requires proper consideration and justification. Otherwise, we open the door to arbitrary decisions at any point during budget execution: a manager in a line ministry could simply decide that they disagree with the government’s policy and move funds from one area to another because they think this is “better.”

Improving credibility

Obviously, there are some technical aspects of the problem that can be addressed, such as improving projections. But in many cases, poor projections are incentivized. What do I mean? In Mexico, the government consistently underestimates revenue to create a surplus that can be distributed discretionally. The problem, which has grown over time, is not primarily technical, although of course it occurs because of a permissive legal and oversight context. You can read more about this [here](#).

In some cases, there are capacity challenges in ministries that can be addressed. Broadly, there may be a tendency to fail, on the part of both legislatures and executives, to carry out sufficient due diligence on capital projects.

A substantial part of the problem has to be addressed through greater visibility and engagement with oversight and the public. Without sufficient eyes on these problems, or demands from outside of government, the incentives for credible budgets remain too low. This is why we believe the democratic perspective on budget credibility is vitally important.