THE UNIVERSITY OF WARWICK

September 2011 No.51

How to Waste a Crisis: Budget Cuts and Public Service Reform

David Hugh Jones
University of Warwick

WORKING PAPER SERIES

Centre for Competitive Advantage in the Global Economy

Department of Economics

CAGE Competitive Advantage in the Global Economy

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

How to Waste a Crisis: Budget Cuts and Public Service Reform

David Hugh-Jones*

07/27/2011

Abstract

In the aftermath of the financial crisis, governments have proposed saving money by reforming public services. This paper argues that tight budget constraints make reform harder. Governments are uncertain which departments are effective. Normally, effective departments can be identified by increasing their budget, since they can use the increase to produce more than ineffective departments. When budgets must be cut, however, ineffective departments can mimic effective ones by reducing their output. Budget cuts thus harm both short-run productive efficiency, and long-run allocative efficiency. These predictions are confirmed in a panel of US libraries. Low marginal productivity libraries reduce output by more than expected in response to a budget cut, and budget setters respond less to observed short-run output elasticity after cutback years.

*Centre for Competitive Advantage in the Global Economy, University of Warwick, CV4 7AL. Tel: +44 (0) 2476 151175. Email: D.Hugh-Jones@warwick.ac.uk. The author thanks Sharun Mukand, Dennis Leach, Hugh Ward and participants at the EPCS and EPSA 2011 conferences.

Keywords: bureaucracy, reform, signaling

JEL code: D73, H12, H61

As government borrowing ballooned in the wake of the 2008 financial crisis, an era of deficit reduction and tight budget constraints beckoned. But political figures around the world looked for the silver lining. Rahm Emmanuel remarked: “You never want a serious crisis to go to waste” (Wall Street Journal, 2008). Hillary Clinton told young Europeans: “Never waste a good crisis” (Independent, 2009). And Barack Obama exhorted the American nation “to discover great opportunity in the midst of great crisis” (USA Today, 2009). On the other side of the Atlantic, management consultants urged the public sector to “us[e] the downturn as a catalyst” to “embrace innovation and rethink delivery models” (Deloitte, 2009), while David Cameron described public service modernization as “not an alternative to dealing with our debts – it’s a key part of it” (Cameron, 2011).

These inspiring slogans are capable of multiple interpretations, but one central one, best expressed by the Cameron quote, is that periods of austerity are the right times for reform. Ordinarily, politicians face few incentives to examine public sector budgets for waste and inefficiency. When budgets are constrained, however, they need to inspect the books more closely, trim dead wood programs, and emerge with a slimmer but more effective government.

This paper explores these issues, with the aim of extending our comprehension of the relationship between budget constraints and efficient allocations. It gives one reason why the hopes expressed above may not be borne out. In the model presented here, savings do not, in general, promote efficiency. Instead, there is a trade-off between savings and efficiency, which bites particularly hard when budgets are tight. The logic is one of signalling. Governments are unsure *a priori* which departments are spending public money effectively. In good times, a government can identify effective departments by increasing their budget on a trial basis. Efficient departments will then be able to do more than inefficient ones with the extra money. In bad times, however, the government cannot afford to expand the budget. It would like to cut the budget of ineffective but over-resourced departments, whose output will be least affected by the cuts. But such departments cannot be identified by trial cuts, because by working inefficiently, they can make it appear that any cut is very harmful.

The argument is illustrated below in Figure 1. The *x* axis measures a bureaucratic department's budget. The *y* axis measures the department's output – a (presumably public) good such as miles of roads maintained, library books lent out, or heart operations performed. A government in office observes a single point in these two dimensions: the current budget (SQ) and the current output. From the government's point of view, this is the point where a set of counterfactual lines cross. In other words, decision-makers must answer the question: "what would happen if we raised, or cut, this department's budget?" Correspondingly, the bureaucratic departments is one of two possible kinds.[1] It may be a high marginal productivity type ("high type" for short), whose output will be greatly increased by extra money, and be seriously harmed by cuts. Or it may be a "low" marginal productivity type, which will gain less from extra cash and be harmed less by cuts. Departmental productivity could differ in this way for many reasons. Departments may vary in their efficiency, and face tasks and external environments

[1] The simplifying assumption of two types is used throughout; the logic extends simply to a continuum of types.

of varying difficulty. Or, the social value of the task they do may be uncertain; some aspects of the task are highly valuable, others less so, and it is hard to untangle these. Since the government may not be able to observe all these factors, it may be unsure of the effects of changes to the budget.

Ex ante, there is no reason that the potential output curves of the two types should cross precisely at the current budget. Instead, we can presume that there is a larger set of possible types; the observation of the status quo budget and output then rules out all but those types whose budget lines pass through that point. Thus, I assume that the government does have reasonable information about the department's current performance. So, for example, the government may know how many miles of road were built in the current year, but be less sure about how many miles could be built with a 10% increase or cut in the budget: exactly how much can costs be driven down in negotiations with contractors? Or, reliable crime figures may be recorded, but it may be unclear how they would be affected by an increase or cut in the budget for policing.

These assumptions exclude, on the one hand, bureaucratic activities whose value is intrinsically hard to measure, such as public funding of the arts, or activities whose outputs are not immediately experienced, such as children's education; and on the other hand, activities with clear outputs that are also offered in competitive markets at public prices, such as perhaps the purchase of IT equipment or stationery.

In forming estimates of potential outputs for different budgets, governments may have more history to go on than a single data point. However, since bureaucratic capacities, external conditions and the social value of outputs all change over time, the government will still face some uncertainty over the bureaucracy's production possibility frontier; this uncertainty will be lowest for relatively recent data; and, since most budget changes are incremental, this data will probably involve budget inputs close to the current one. For simplicity, I therefore assume that output at the current budget is known for sure, with uncertainty increasing as we move farther from this point.

Suppose that the long-run cost of the government's funds is given as in the curved line in the picture. Then, the optimal budget for a high type department would be at point B, where marginal benefit is equal to marginal cost – an increase on the current budget. The optimal budget for a low type department would be a cut to point A.

One way to find out the quality of the department would be to increase the budget for a short time, for example to point B itself. Suppose that the bureaucracy is a single actor, who can costlessly produce any output up to the line for the department's type, and who is solely interested in maximizing the departmental budget *à la* Niskanen (1971).[2] With the extra cash, a high type department can produce strictly higher output than the low type department could. By doing so, the bureaucrat can demonstrate the department's competence and continue to receive a higher budget in future. A low type department, on the other hand, can only produce up to its output line. After the trial period, efficient long-term budget allocations can be made, according to the department's type, at either A or B.

Now, however, suppose that the government faces a tight short-run budget constraint. For example, it may have entered office with high government borrowing

[2]This is a simplification (Dunleavy, 1985), but it may not be an unrealistic one if interpreted within the context of protecting an existing budget from cuts.

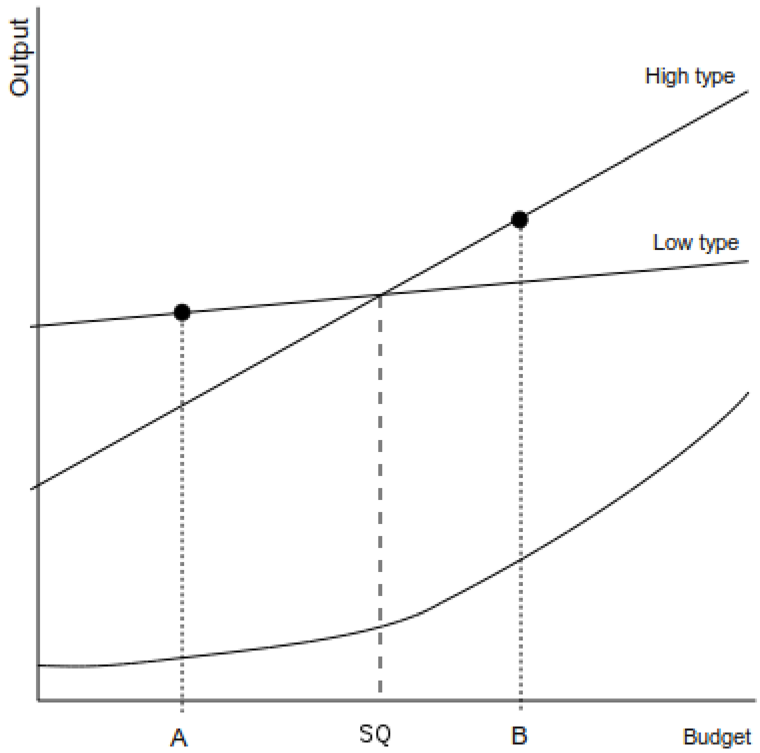


Figure 1: Bureaucratic departments

left over from the previous administration, and bond markets nervous about its ability to repay. Or, a particular minister may have been allocated a low budget by the Prime Minister. Say that the maximum budget available for the trial period is below the status quo budget SQ.

The situation is now very different, because a high type department can no longer perform better than a low type department could. Since the high type's marginal productivity at current output is higher, it will be more affected by the budget cut. Even when the high type performs at its maximum, the low type can match it by costlessly lowering its output. In other words, over-resourced bureaucrats can claim that they are more harmed by budget cuts than they really are, simply by reducing their productivity. As a result, in the short run there will be productive inefficiency due to destroyed output; in the long run, departments cannot be distinguished, and there will be allocational inefficiency.

The idea that bureaucrats may respond to budget cuts by exaggerating, or even deliberately aggravating, their effects is part of the folk wisdom of public administration. The tactic is known as "sore thumbs", "bleeding stumps", or in the US as the Washington Monument Ploy, named after the US Park Service's regular threats to close the Washington Monument if their budget were to be cut. Wildavsky (1979 [1964] p.102) mentions "cut the popular program" as a strategy for agencies seeking to reverse budget cuts. Several authors in Hood and Wright (1982) describe the bleeding stumps tactic; as the editors put it, "the course of retrenchment is fatally distorted by bureaucratic preferences". The current round of post-crisis budgets has led to other examples. In 2008, California governor Arnold Schwarzenegger rhetorically proposed a budget which included the early release of 37,000 prisoners. Eric Pickles (the current UK Secretary of State for Communities and Local Government) has accused Labour local authorities of operating a "bleeding stumps strategy" (BBC, 2011). However, the distortions of retrenchment go beyond the direct effects of shirking by ill-motivated bureaucracies. At least as important is the *information loss* because the government – the principal in this situation – cannot identify the truly essential programs, where cuts will seriously harm the provision of public services. While the loss from shirking is immediate, the information loss is more long-term and may outlast the crisis period. And, of course, the information loss occurs even when the bureaucracy is not shirking at all, but is genuinely damaged by cuts. The theory here also clarifies why stories of "bleeding stumps" emerge in times of cutbacks and not of growth: the Washington Monument ploy can be used when the government must cut spending, but there is no corrresponding tactic for when budgets are going up.

Below the model is developed in detail. The empirics examine a panel of US libraries. The data bear out the model's prediction that low marginal productivity bureaucracies will be more likely to underperform in years when their budget is cut, while high marginal productivity bureaucracies will be less affected. They also support the notion of long-run allocational inefficiency: decision-makers' budget allocations are less sensitive to observed bureaucratic performance after a budget cut.

Model

There is a single bureaucracy, which is either a high-productivity type, or a low-productivity type. Given a budget of x, the high type can produce output to a maximum value of $\alpha_H + \beta_H x$, while the low type can produce $\alpha_L + \beta_L x$ with $\beta_L < \beta_H$.[3] This output level can be thought of as the result of an underlying decision problem among the bureau's personnel. For instance, output beyond the default level (α_H or α_L) might be produced by hiring homogenous workers at a unit cost of 1, who produce at β_H or β_L depending on the efficiency of the department's technology. Or, different types of bureaucrats may trade off work and on-the-job leisure differently. However, the bureaucracy may also make the political choice to produce output lower than the maximum. For example, the bureaucracy's manager may allocate money inefficiently so that output is lost. Producing output below the maximum has a positive cost, since it involves (for example) reorganizing work routines so as to be less efficient, or consuming more leisure than is optimal. I assume for simplicity that this cost is small.[4]

Current spending is at a default level $\bar{x}$ and output is $\bar{y}$, where $\bar{y} = \alpha_H + \beta_H \bar{x} = \alpha_L + \beta_L \bar{x}$ is the point at which both output lines cross. As argued above, this is not a coincidence: the government's observation of current productivity can be thought of as reducing its uncertainty over the bureaucracy's output function, perhaps from a much larger set of types.

Rather than imposing a hard budget constraint as above, I assume the government faces a short-term cost of funds $\hat{c}(x) = kc(x)$. Marginal cost is increasing, since the government's credit in the markets is not unlimited, and since there are com-

[3]The linear functional form is not essential: concave output and a weakly convex cost of funds would give similar results. The important condition is that the types' output curves cross only once, at the status quo point.

[4]Relaxing this assumption would have predictable effects. As the cost of producing below the optimum grows, it becomes harder for the low type bureaucracy to pool in the case of budget cuts. Since this cost is difficult to observe empirically, introducing it as a parameter would add little to the model's explanatory value.

peting spending priorities. The parameter k reflects how these conditions may vary. When there are many competing priorities and funds are tight, or when the government must pay high interest to borrow, k will be high. I assume $c', c'' > 0$, $c(0) = 0$ and $c'(x) \to 0$ as $x \to 0$. The long-term cost of funds is just $c(x)$.

The timing is as follows:

1. Nature draws the bureaucracy's type $\tau \in \{H, L\}$ which is high (H) with probability π.

2. The government chooses a first-period budget x_1 at a cost $kc(x_1)$.

3. The bureaucracy chooses a level of output $y_1 \leq \alpha_\tau + \beta_\tau x_1$. Inefficient levels of output $y_1 < \alpha_\tau + \beta_\tau x_1$ may be achieved by, for example, allocating funds to inefficient uses.

4. The government observes y_1 and chooses a second-period budget x_2 at a cost $c(x_2)$.

5. The bureaucracy produces $y_2 = \alpha_\tau + \beta_\tau x_2$.[5]

6. Payoffs are realised. The bureaucracy receives $\delta x_1 + x_2$ and the government's utility is $\delta[y_1 - kc(x_1)] + y_2 - c(x_2)$. Here δ is a parameter reflecting the relative length of the first period, in other words the time until output becomes accurately measurable. This time may be greater for some outputs than for others. [6]

[5]Second period productive efficiency is guaranteed by the bureaucracy's small cost of destroying output.

[6]For instance, the effectiveness of infrastructure investment may take longer to become clear than the effectiveness of hardship payments to Old Age Pensioners. Recent British governments have invested extensively in performance management statistics for local authorities, health and education.

We first look for conditions in which the government can distinguish the types after the first period. If this is so, the second period budget will be x_H solving $c'(x_H) = \beta_H$ for a high type and x_L solving $c'(x_L) = \beta_L$ for a low type. Since $x_H > x_L$, each type of bureaucracy would prefer to appear like the high type. For $x_1 \leq \bar{x}$, because the low type's maximum possible output exceeds the high type's, the low type can match any output that the high type could produce. Thus it must be that $x_1 > \bar{x}$. Then the high type bureaucracy can exceed any possible output of the low type, and so identify itself to the government, by producing efficiently.

So for a high enough first period budget, types can always be distinguished. However, if the short-run cost of funds is large, the government may prefer not to do so. The short-run optimum budget allocation, ignoring the second period and assuming that both types of bureaucrat produce efficiently, would be x_1^* solving

$$c'(x_1^*) = \frac{\pi\beta_H + (1-\pi)\beta_L}{k}. \tag{1}$$

If $x_1^* > \bar{x}$ then the short-run optimum will also allow the government to distinguish the types. But if k is large enough this will no longer hold. The government then faces a choice: keep the budget high in the first period in order to observe the bureaucracy's type, or spend less and fail to do so.

Keeping the budget high requires allocating $x_1 = \bar{x}$ in the first period.[7] The gov-

[7] In fact, to learn the department's type we require $x_1 > \bar{x}$. This open set has no minimum, so the government's optimal choice may not be well-defined. This is a purely technical point, and I simply assume that the government must pay some arbitrarily small extra amount above $\bar{x}$ to differentiate the types. An alternative fix would be to add some noise to the department's output. In this case, larger budget increments above $\bar{x}$ would give continuously more accurate signals of deparmental type, and the optimal budget would trade off signal accuracy against period 1 optimality, typically coming in strictly above $\bar{x}$. This suggests an information-based rationale for cycles in spending levels, but the logic is not pursued further here.

ernment's expected total payoff is then

$$\delta[\bar{y} - kc(\bar{x})] + \pi(y_H - c(x_H)) + (1 - \pi)(y_L - c(x_L)) \tag{2}$$

where $y_\tau = \alpha_\tau + \beta_\tau x_\tau$ is type τ's output after an optimal choice of period 2 budget.

Alternatively, the government could spend less than $\bar{x}$. If so, then in equilibrium both types will produce the same output. For, if not, the low type could pool with the high type by changing output, and would then increase its period 2 budget. I assume that for $x_1 < \bar{x}$, the high type produces efficiently and the low type matches it. As a result the government's total payoff will be

$$\delta[\alpha_H + \beta_H x_1 - kc(x_1)] + \pi(\alpha_H + \beta_H x_2^*) + (1 - \pi)(\alpha_L + \beta_L x_2^*) - c(x_2^*) \tag{3}$$

where x_2^* solves the second period first order condition

$$c'(x_2^*) = \pi\beta_H + (1 - \pi)\beta_L. \tag{4}$$

Optimizing over x_1, observe that the government should treat the bureaucracy like a high type, since the low type will pool at the same output. Thus the optimal choice $\hat{x}_1$ solves $c'(\hat{x}_1) = \beta_H / k$.

Comparing the two alternatives, the net benefit of choosing $\bar{x}$ can be split into two components. There is a period 1 negative gain of

$$\delta[\beta_H(\bar{x} - \hat{x}_1) - k(c(\bar{x}) - c(\hat{x}_1))]. \tag{5}$$

This is negative, and decreasing (without bound) in k and δ.[8] On the other hand there is a period 2 gain from knowing the types, of

$$\pi[\beta_H(x_H - x_2^*) - (c(x_H) - c(x_2^*))] + (1-\pi)[\beta_L(x_L - x_2^*) - (c(x_L) - c(x_2^*))] \quad (6)$$

This is positive and unrelated to δ or k. Summing these gains, if δ or k is high enough, then the government will strictly prefer to choose $\hat{x}_1$ in period 1, enacting a budget cut which weakens its ability to distinguish an effective from an ineffective department.

Empirics: behaviour of the bureaucracy

One implication of this theory is that governments' budget allocations do not decrease continuously as the cost of funds k increases. Instead, at first the government prefers to keep funding steady, so as to learn about the bureaucracy; a further increase in k leads to a discontinuous drop in the budget when maintaining the existing level can no longer be afforded. Another implication is that productivity decreases after budget cuts. However, both these predictions are already made by extant theories: budget cuts may harm bureaucratic morale (Hood and Wright, 1982); budget-setters may use the status quo as a reference point (Wildavsky, 1964); and so forth. A more specific implication of the model is that high- and low-productivity departments will be affected differently by budget cuts. High-productivity departments will produce no less output after a budget cut than in normal times. Low-productivity departments will produce less, since they will

[8] Shown in the Appendix.

N	91687			
# unique library systems	6695			
	Min	Median	Mean	Max
# years data per library	1	12	9.2	17
Income ($)	504	369,900	1,410,000	249,700,000
% change income	-99.7%	5.0%	8.5%	114.3%
Attendance	1	68,790	211,800	17,940,000
Circulation	0	102,400	329,200	23,240,000
Public Service Hours	0	2,938	5,583	250,400

Rows with missing data were excluded.

Table 1: US Public Libraries Survey, descriptive statistics

pool with high-productivity departments.

To test this prediction, I examine data from the US Public Libraries Survey, covering 9965 US public library systems[9] over the time period 1993 to 2009. Libraries are a plausible case for the theory. First, they produce measurable outputs – books lent, total visits and inter-library loans serviced – which are recorded in the survey data. Second, the "bleeding stumps" tactic appears to be known in the sector: library management textbooks mention it as a potential method of avoiding budget cuts (e.g. Wood and Young, 1988). Table 1 shows descriptive statistics for this dataset.

The Public Libraries Surveys include budget data, and also data on various outputs, including attendance, circulation, and total annual Public Service Hours (summed over all library branches). I use Public Service Hours (PSH) as the main measure of output, since these should be least affected by variation in consumer demand.

To get an intuition for the empirics, consider Figure 2. Each figure plots logged

[9] A public library system is a single managerial and accounting unit, which may comprise one or more physical branches.

yearly PSH against logged yearly budget for a single library in the dataset. The straight line is the result of a regression using only those years when the library's (nominal) budget increased. Triangles show years when the budget decreased. For Viking Library System Operations, Minnesota, the line is relatively steep: in the model's terminology, this library could be a "high type". Correspondingly, in 3 out of 4 years of budget cuts, attendance was higher than the regression predicts. For Ypsilanti District Library, Michigan, by contrast, the line is not merely shallow but negatively sloped: the library achieved fewer Public Service Hours when its budget was larger. This is a "low type". In 3 out of 4 years when its budget was cut, the Ypsilanti library performed worse than the prediction from the regression. As the model predicts, the less productive library underperformed when the budget was cut; the more productive library did not.

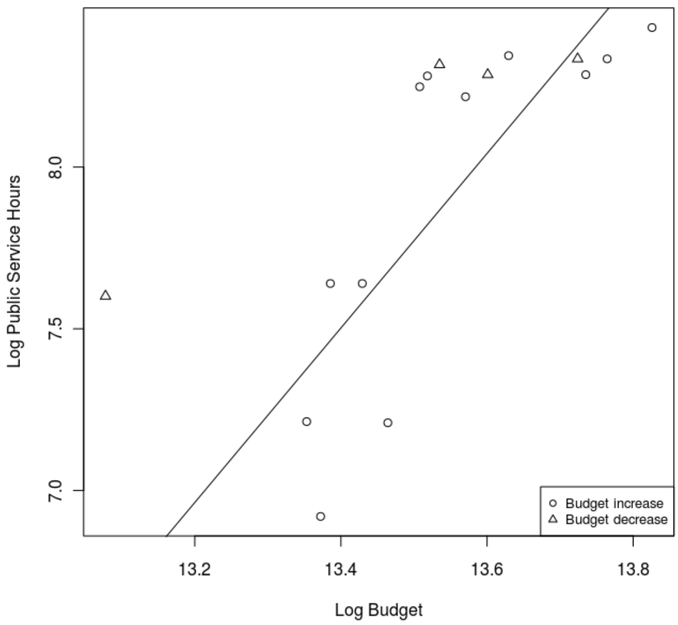
To repeat this analysis for the entire dataset,[10] I ran the following procedure for each library:

1. Regress log PSH on log income, using only years in which income increased.

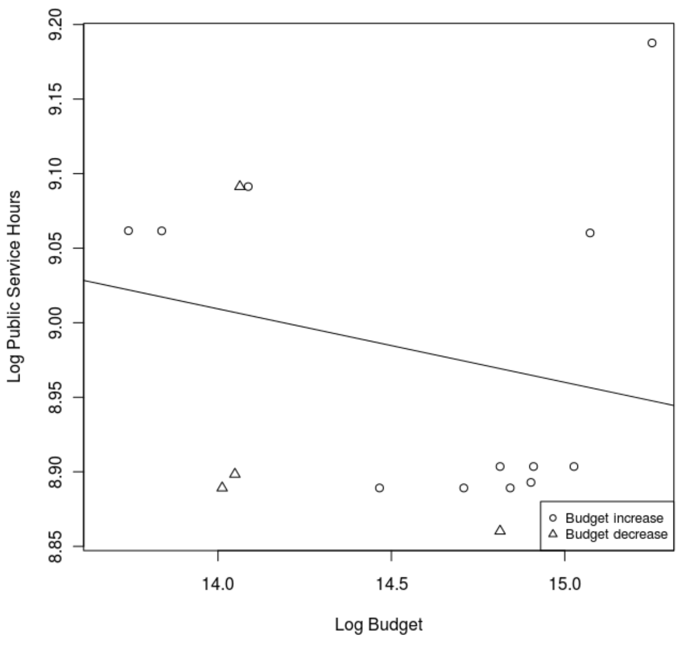
2. Examine the difference between real PSH and PSH predicted from this regression (i.e. the residuals), for all "cutback years" in which income decreased.

3. Record the slope of PSH on income, and the proportion of residuals which were negative.

[10] I used every library with at least 12 years of data during which the budget increased, and at least one year during which the budget decreased.



(a) Viking Library System Operations, Minnesota



(b) Ypsilanti District Library, Michigan

Figure 2: Public Service Hours versus budget, for two libraries

My hypothesis is that *in cutback years, low marginal productivity libraries will be more likely to have lower-than-predicted public service hours,* where marginal productivity is measured by the slope of PSH on library income.

Figure 3 plots marginal productivity (i.e. the slope of the PSH-income regression) against the proportion of cutback years in which actual hours were fewer than predicted (ie of cutback years with negative residuals). As predicted, the more productive libraries, with the higher slopes, had fewer negative residuals in years of budget decrease (Kendall's test of correlation: $p < 0.001$). Figure 4 shows the same data aggregated by deciles of productivity. The size of the effect is quite substantial: the least productive libraries underperformed in around 60% of cutback years, while the most productive libraries underperformed in around 40% of cutback years.

As a robustness check I repeated the analysis using a quadratic in log income, measuring the PSH-income slope at the mean of the library's log income. Results were substantively unchanged (Kendall's test: $p < 0.001$). I also reran the analysis using all libraries with at least 6 years when the budget increased, and again results were robust (Kendall's test: $p < 0.001$). Lastly I repeated the basic analysis on subsets of the data. The relationship between productivity and underperformance during cutbacks was negative in 39 out of the 47 states for which enough measurements were available. It also held in every decile of library size, as measured by total income in 2000, was significant at $p < 0.05$ (Kendall's test) in 9 out of 10 deciles, and was always significant at $p < 0.10$.

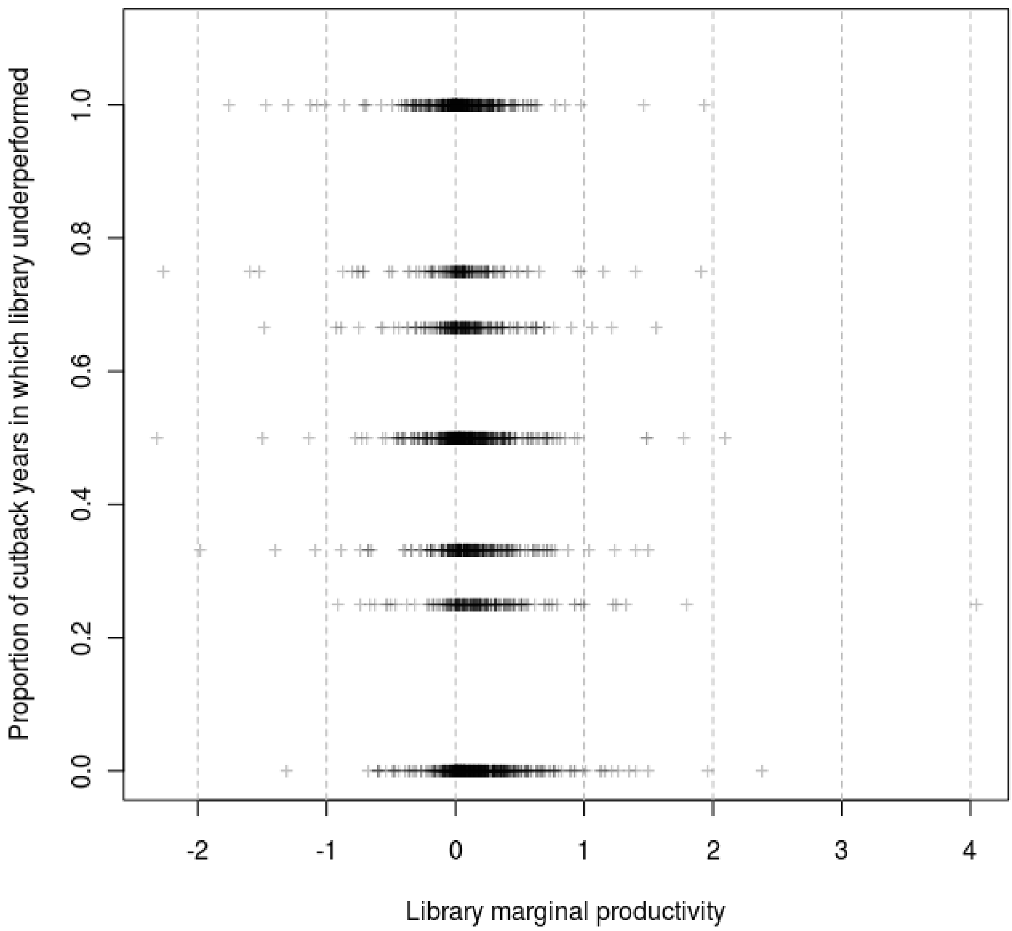


Figure 3: Library productivity versus performance during cutbacks

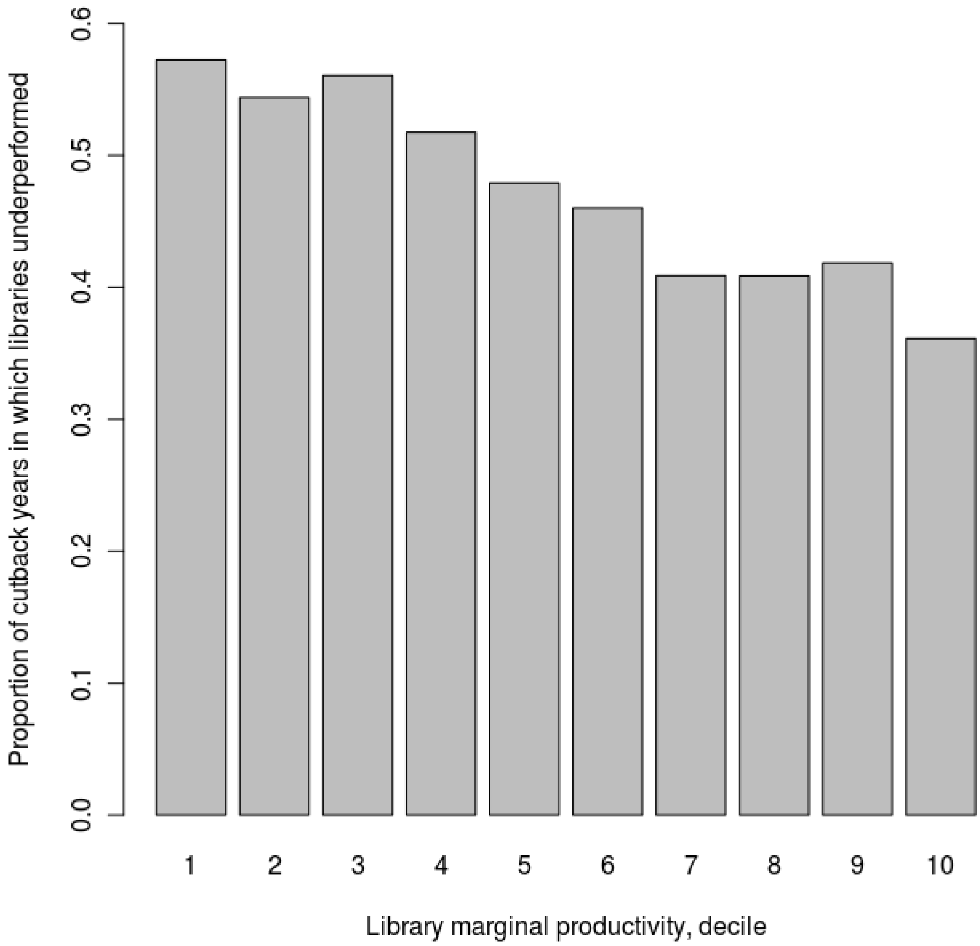


Figure 4: Library productivity versus performance during cutbacks: by decile

Empirics: behaviour of the budget-setter

A further prediction of the model focuses on the government's behaviour. In normal times, the department's first period output informs the government of the department's type, and the government allocates a larger budget to more productive departments. After cutbacks, however, first period output is no longer informative and the government ignores it in setting the second period budget. Thus, cutbacks reduce allocational efficiency in the longer term.

The US libraries in the PLS dataset receive on average about 77% of their income from local governments, with about 10% coming from state governments. I examine how *budget change*, the percentage change in a library's income over years t and $t-1$, is affected by its short-run output *elasticity* in the previous two years, defined as

$$\frac{\text{\% change in total service hours}}{\text{\% change in budget}},$$

over years $t-1$ and $t-2$. *Elasticity* will be high if the library increases total service hours in response to a budget increase, or if it decreases total service hours in response to a budget cut: in other words it is a short-run estimate of marginal productivity (in the model's terms, of the library's type). If governments indeed learn more from, and react more to, recent performance during normal years than cutback years, then the coefficient of *elasticity* upon *budget change* will be larger during normal years. Table 2 shows the results.

The first column provides a basic check that budget-setters respond to a library's short-run elasticity. A one standard deviation increase in *elasticity* (by 0.46) is

	(1)	(2)	(3)	(4)	(5)	(6)
DV	Total income			Income from local government		
Intercept	4.8 (0.039)***	4.4 (0.048)***	3.6 (0.58)***	5.1 (0.039)***	5.1 (0.045)***	4.5 (0.51)***
Elasticity	0.28 (0.079)***	0.60 (0.090)***	0.54 (0.095)***	0.18 (0.074)*	0.40 (0.08)***	0.35 (0.086)***
Cutback $t-1$	–	1.6 (0.098)***	2.0 (0.09)***	–	0.15 (0.090)	0.48 (0.083) ***
Elasticity $\times$ cutback $t-1$	–	-1.1 (0.19)***	-1.1 (0.19)***	–	-0.85 (0.18)***	-0.79 (0.17)***
State FE	No	No	Yes	No	No	Yes
Year FE	No	No	Yes	No	No	Yes
S.E.s	Clustered	Clustered	Independent	Clustered	Clustered	Independent
N	65309	65309	65309	63748	63748	63748
# unique libraries	6125	6125	6125	5975	5975	5975
R^2	0.00015	0.0051	0.028	0.000074	0.00047	0.022

Notes: S.e.s in columns 1, 2, 4 and 5 are clustered by library. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.
Outlier library-years, in which the library received a more than 33% budget increase or budget cut, or library-years with an elasticity of absolute value more than 2 are excluded.

Table 2: How library budget-setters respond to performance

associated with an approximately one-and-a-third standard deviations change to the budget increment. Columns 2-3 include a dummy variable for a cutback year in $t-1$, and interact this with elasticity measure. After cutback years, the effect of elasticity is significantly less than after a normal year, as predicted by the theory; indeed, it is significantly negative, so that when an initial cut has a large effect on output, then there is a *smaller* subsequent budget increase on average. The third column adds state and year fixed effects.[11] Columns 4-6 repeat the analysis, but use percentage change in local government funding as the dependent variable, with broadly similar results.[12]

Extension: monitoring the bureaucracy

When the cost of funds are high enough to rule out temporary spending increases, governments may seek alternative ways to control the bureaucracy. To see this, suppose that by paying a monitoring cost m, the government can ensure that the bureaucracy produces at its efficient level $y = \alpha_\tau + \beta_\tau x$ in period 1. For instance, certain budget items may be ring-fenced, harming the bureaucracy's flexibility in responding to changing circumstances but simultaneously forcing it to cut the fat, not the muscle. Or, *ex post* checks may be used to discover and sanction inefficient spending patterns. This will be unnecessary when $x_1 \geq \bar{x}$, but will be worthwhile for high enough k. For, when both types produce efficiently, there are two benefits – first a period one benefit, reflecting the extra productivity of a low

[11] Other specifications were tried, including per-state time trends, and adding lagged budget change directly as a control. Coefficients on *elasticity* and *elasticity* $\times$ *cutback* $t-1$ remained highly significant in the expected direction. Results are available on request.

[12] The "cutback in $t-1$" dummy is still defined with reference to total funding, since this is what determines the ability of low-type bureaucracies to pool.

type bureaucracy that is prevented from pooling; second, the indirect benefit from learning the types and allocating resources efficiently in period two.

Specifically, after paying the monitoring cost, the government can choose $x_1 = x_1^*$. Each type of bureaucracy then produces efficiently in round 1, and the government can then correctly distinguish the type it faces, resulting in round 2 allocations of x_H or x_L. On the other hand, without monitoring, the government's payoff is given by (3) with x_1 set to $\hat{x}_1$. If the monitoring cost is not too high, and if the credit crisis is sufficiently serious (k is high), then the government will always prefer to monitor.[13]

Conclusion

To enhance our understanding of the dilemmas of spending cuts, this paper developed a simple theory in which government is uncertain about the effect of changing a department's budget. The theory predicts that less efficient departments will be more affected by budget cuts, since they "pool" with more efficient departments so as to exaggerate the effect of their loss. This prediction was confirmed in a sample of US libraries. It was also confirmed that budget-setters are less reactive to short-run estimates of the budget-output slope in the year after a budget cut. These results are compatible with the claim that budget cuts harm both short-run productive efficiency, and long-run allocative efficiency.

It is worth reiterating this theory's domain of application. The argument applies to bureaucracies whose output is measurable in the short term – so, for example, probably not to arts organizations, whose output is not easily measurable, nor to

[13] Proved in the Appendix.

long-term infrastructure projects or investments whose impact is not immediately visible. Also, the assumption of a budget-maximizing bureaucracy implies that bureaucratic rewards cannot be directly linked to performance; in other words, contracts are incomplete. Incomplete contracts are a common theme in political economy. Nevertheless, governments in many developed countries have invested heavily in measuring and rewarding bureaucratic performance, and where this is successful, the information problem will no longer bite. Conversely, the Washington Monument ploy need not be limited to politics; the same behaviour may arise within firms. It would be interesting to find examples.

The practical implication of this paper is straightforward: reforms and cuts do not mix, because cuts exacerbate the informational problems between government and the bureacracy. There may, nevertheless, be other ways in which spending crunches ease the path of public service reform. For example, they may increase electoral support for tough austerity measures. In the coming few years, as Western governments' spending cuts take effect, there should be many opportunities for further research in this area.

Appendix

Proof that (5) **is negative and decreasing in δ and k.**

Write (5) as

$$\delta[\beta_H(\bar{x}-\hat{x}_1)-k\int_{\hat{x}_1}^{\bar{x}}c'(x)dx].$$

By the FOC on $\hat{x}_1$, $c'(\hat{x}_1) = \beta_H/k$ and so by $c'' > 0$, the above is less than

$$\begin{aligned} & \delta[\beta_H(\bar{x}-\hat{x}_1) - k\int_{\hat{x}_1}^{\bar{x}} \frac{\beta_H}{k} dx] \\ = \ & \delta[\beta_H(\bar{x}-\hat{x}_1) - \beta_H(\bar{x}-\hat{x}_1)] \\ = \ & 0. \end{aligned}$$

That (5) decreases without bound in δ is then immediate. To show it decreases in k, suppose $\bar{k} > \underline{k}$. Write $\bar{x}_1$ for the solution to $c'(x) = \beta_H/\bar{k}$, and $\underline{x}_1$ for the solution to $c'(x) = \beta_H/\underline{k}$. Observe that $\bar{x}_1 < \underline{x}_1$Then for $\underline{k}$, (5) is

$$\underline{B} \equiv \delta[\beta_H(\bar{x}-\underline{x}_1) - \underline{k}\int_{\underline{x}_1}^{\bar{x}} c'(x)dx],$$

while for $\bar{k}$ it is

$$\begin{aligned} & \delta[\beta_H(\bar{x}-\bar{x}_1) - \bar{k}\int_{\bar{x}_1}^{\bar{x}} c'(x)dx] \\ = \ & \delta[\beta_H(\bar{x}-\underline{x}_1+\underline{x}_1-\bar{x}_1) - \underline{k}\int_{\underline{x}_1}^{\bar{x}} c'(x)dx - \underline{k}\int_{\bar{x}_1}^{\underline{x}_1} c'(x)dx - (\bar{k}-\underline{k})\int_{\bar{x}_1}^{\bar{x}} c'(x)dx] \\ = \ & \underline{B} + \delta[\beta_H(\underline{x}_1-\bar{x}_1) - \underline{k}\int_{\bar{x}_1}^{\underline{x}_1} c'(x)dx - (\bar{k}-\underline{k})\int_{\bar{x}_1}^{\bar{x}} c'(x)dx]. \end{aligned}$$

Of the terms in square brackets, the first two sum to less than zero since $c'(\underline{x}_1) = \beta_H/\underline{k}$ and $c'' > 0$, and the last term is negative. Lastly observe that as $k \to \infty$, $\hat{x}_1 \to 0$ and (5) therefore approaches $\delta[\beta_H\bar{x} - kc(\bar{x})]$ which decreases towards infinity with k. QED

Proof that when m is not too high and k is sufficiently high, the government will prefer to monitor

Specifically, the condition is $m < (1-\pi)(\alpha_L - \alpha_H)$. The government's total payoff after paying to monitor is:

$$-m+\delta[\pi(\alpha_H+\beta_H x_1^*)+(1-\pi)(\alpha_L+\beta_L x_1^*)-kc(x_1^*)]+\pi(y_H-c(x_H))+(1-\pi)(y_L-c(x_L)). \tag{7}$$

Comparing (3) to (7) shows the government prefers to monitor if

$$m < B_1 + B_2,$$

where B_1 represents the period 1 benefit of monitoring and B_2 the period 2 benefit:

$$\begin{aligned} B_1 &= \delta[\pi(\alpha_H+\beta_H x_1^*)+(1-\pi)(\alpha_L+\beta_L x_1^*)-(\alpha_H+\beta_H \hat{x}_1)-k[c(x_1^*)-c(\hat{x}_1)]]; \\ B_2 &= \pi(\beta_H(x_H-x_2^*)-[c(x_H)-c(x_2^*)])+(1-\pi)(\beta_L(x_L-x_2^*)-[c(x_L)-c(x_2^*)]). \end{aligned}$$

B_2 is easily seen to be positive since x_H and x_L are optimal for a high and low type respectively. B_1 can be split into two components. First there is the value from preventing pooling, thus increasing the productivity of the low type. Second there is the benefit of being able to choose an *ex ante* optimal budget. Thus

$$\begin{aligned} B_1 &= \delta[\pi(\alpha_H+\beta_H x_1^*)+(1-\pi)(\alpha_L+\beta_L x_1^*)-kc(x_1^*) \\ &\quad -\{\pi(\alpha_H+\beta_H\hat{x}_1)+(1-\pi)(\alpha_L+\beta_L\hat{x}_1)-kc(\hat{x}_1)\} \\ &\quad +\{\pi(\alpha_H+\beta_H\hat{x}_1)+(1-\pi)(\alpha_L+\beta_L\hat{x}_1)-kc(\hat{x}_1)\} \\ &\quad -(\alpha_H+\beta_H\hat{x}_1-kc(\hat{x}_1)] \end{aligned}$$

The first two lines are the benefit of choosing an optimal x_1. They sum to a positive amount since x_1^* is the optimal choice when types are unknown, and therefore maximizes $\pi(\alpha_H+\beta_H x)+(1-\pi)(\alpha_L+\beta_L x)-kc(x)$.

The last two lines are the benefit of preventing pooling. They simplify to

$$(1-\pi)(\alpha_L-\alpha_H+(\beta_L-\beta_H)\hat{x}_1)$$

Now, as k grows large $\hat{x}_1 \to 0$ and this term approaches $(1-\pi)(\alpha_L-\alpha_H)$. Thus if $m<(1-\pi)(\alpha_L-\alpha_H)$, for k large enough the government will find it worthwhile to monitor.

References

BBC. 2011. “Pickles in ’bleeding stump’ claim.” *BBC* .
URL: *http://www.bbc.co.uk/news/uk-politics-12657517*

Cameron, David. 2011. “Speech on public service reform.”.

Deloitte. 2009. Turning the Tide: Opportunities for Public Sector organisations in an economic downturn. Technical report.

Dunleavy, Patrick. 1985. “Bureaucrats, Budgets and the Growth of the State: Reconstructing an Instrumental Model.” *British Journal of Political Science* 15(3):299–328.

Hood, C. and M. Wright. 1982. “Big government in hard times.” *Public administration* p. 109.

Independent. 2009. “Clinton: ’Never waste a good crisis’.” *The Independent* .

Niskanen, W. A. 1971. “Bureaucracy and representative government.”.

USA Today. 2009. “Obama: Crisis is time of ’great opportunity’.” *USA Today* .

Wall Street Journal. 2008. “In Crisis, Opportunity for Obama.” *Wall Street Journal* .

URL: *http://online.wsj.com/article/SB122721278056345271.html*

Wildavsky, A. 1964. *The politics of the budgetary process*. Little, Brown and Co.

Wood, Elizabeth J. and Victoria L. Young. 1988. *Strategic marketing for libraries*. Greenwood Publishing Group.