Optimal Rollover Policy with Multi-Period Budgets *

Julian Koller †

December 26, 2024

Abstract

In many organizations, agents operate on a fixed budget and allocate funds across sub-periods of the fiscal year. Commonly, unspent budget must be returned to the principal at year-end to prevent policy drifts. However, this savings constraint may induce inefficient spending patterns, such as expenditure surges before the budget expires. This study characterizes optimal budget rollover policy in a model which captures this trade-off. Applying it to Swiss federal consulting spending, I show that substantial welfare gains are possible by allowing agencies to retain one third of unspent funds. I verify the model's predictions exploiting a staggered reform liberalizing rollover.

Keywords: Public Finance, Rollover, Budget Rules

*Insert acknowledgements if accepted
†ETH Zürich, KOF Swiss Economic Institute, Leonhardstrasse 21, 8092 Zürich

1 Introduction

The decision to spend or to save subject to a budget constraint and uninsured idiosyncratic risk lies at the very heart of economics. However, relatively little attention has been given to the case in which a borrowing-constrained agent faces a fixed budget over multiple periods, allowing costless saving, but ending with partial rollover of unspent funds to the next budgetary cycle. This structure of the decision problem is very common in large organizations such as big corporations, public bureaucracy, and also universities. Many of these organizations worry that allowing agencies unrestricted saving at the end of the fiscal year can lead to a loss of control over spending policies (i.e., a "policy drift"). This may be the case if the agent, who implements the spending sequence, has incentives that do not perfectly align with those of the principal, who allocates the budget. On the other hand, concerns persist that forcing the agent to return all unspent budget by the end of the budgetary cycle ("use-it-or-lose-it") leads to an inefficient distribution of spending throughout the year, most notably incentivizing large spending spikes just before the budget expires. Figure 1 illustrates this pattern for federal contract obligations in the US, highlighting September, the final month of the fiscal year, which shows a pronounced spending spike before the agencies' budgets expire.

Is this pattern inefficient, and if so, can budget rollover rules improve it? A small but growing body of research, as reviewed below, concludes that the answer to the first question is affirmative. Regarding the second question, despite prominent debates among practitioners (see, for example, General Accounting Office 1980 or General Accounting Office 1998), there is surprisingly little economic research focused on addressing it. Yet, the numerous changes in public finance rollover regimes in different countries underline both the relevance of such a policy and the uncertainty which surrounds it.[1]

Taking a step towards a better understanding of such reforms, this paper aims to theoretically characterize the trade-offs involved in designing year-end rollover rules for agencies operating on a fixed-period budget which is allocated by a principal. Importantly, my setup allows for a wedge between the agency's and the principal's flow utility. Moreover, my approach is going to account for the possibility of budgetary slack, i.e., systematic non-exhaustion of budget. Finding the welfare-maximizing rollover rule for the principal in such a setup is of great importance to many organizations featuring agencies on fixed-period budgets. I apply the model to do this for the case of Swiss public consulting spending, illustrating that substantial welfare gains relative to use-it-or-lose-it are possible by optimizing rollover policy. To that end, I generalize the model from Liebman and Mahoney (2017), also building on the extensions proposed by Baumann (2019). Each of

[1] To name some examples, the UK switched from a use-it-or-lose-it to a more liberal rollover regime in 1998, to then return to a more constrained rule again in 2001 (Crawford et al., 2009; Baumann, 2019). The US generally operates under a no-rollover rule, but granted limited rollover to the Department of Justice since the 1990s. Moreover, Washington and Oklahoma allow partial rollover on the state level (McPherson, 2007). Another case is Switzerland, which liberalized rollover in 2017, as will be discussed in detail in Section 3.4.

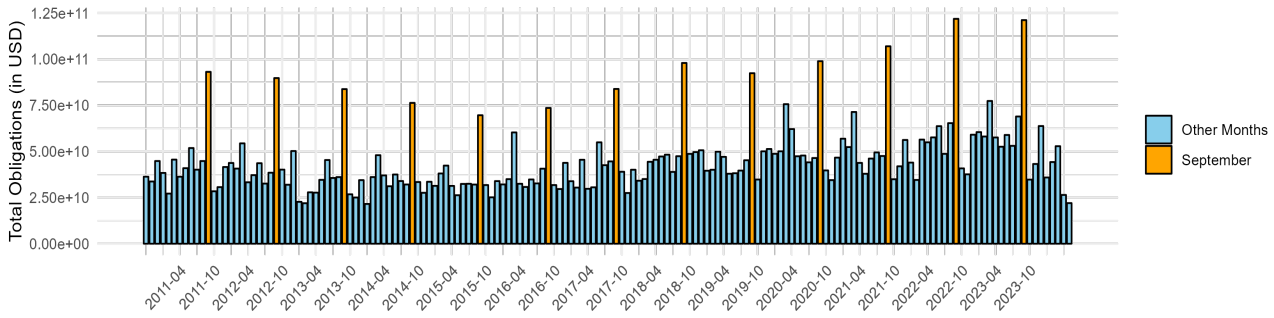


Figure 1: Total US federal contract obligations from 2011 to 2024. The highlighted bars represent September, the last month of the US fiscal year, during which contract spending soars to about twice its average value throughout the rest of the budgetary period. Data source: usaspending.gov.

these models highlights one of the channels which are considered to be the main drivers of uneven spending throughout the fiscal year: Precautionary savings and agency misalignment. In the former, year-end spending spikes originate from disbursed buffer stock savings and in the latter they are driven by policy drifts of the agency, which are heightened towards the end of the fiscal year. To the best of my knowledge, this paper is the first to analyze budget rollover policy while simultaneously considering both of these channels, which turns out to be essential for matching agency spending behavior.[2]

An initial and important insight from the model is that heightened year-end spending is not necessarily indicative of inefficiency by the agent. In fact, also the principal would not prefer smooth spending in expectation throughout the fiscal year. Intuitively, since the principal is borrowing-constrained, spending a large share of the budget in the early months of the fiscal year is more risky, as there is a higher probability to hit the liquidity constraint before the new budget arrives. Thus, delaying spending to some degree is a natural property of the optimal policy for a risk averse principal in the face of uncertainty. Therefore, observing a spending spike as in Figure 1 is not a sufficient statistic for concluding that this spending pattern reflects sub-optimal agency behavior. Instead, full counterfactual simulations are required to evaluate the welfare implications of liberalizing rollover. Therefore, I calibrate the model to sub-annual spending sequences of Swiss public bureaucracies' consulting expenses, as the institutional context in this case closely resembles the setup of the model. This is important, as I argue below that calibrations to a spending series that only reflects part of the expenses within the relevant budget can produce misleading results. Furthermore, I derive how to value a sequence of spent and returned budget by the agency from the perspective of the principal and use this method to evaluate counterfactual rollover rules. This allows me to quantitatively pin down optimal rollover policy from the principal's standpoint. Even though the calibration is of course context-specific, the

[2]Consider Section 3 for details.

relevant trade-offs are general and the framework can be used to evaluate optimal rollover rules in other settings. Subsequently, I analyze a reform liberalizing rollover in Switzerland to support the structural counterfactuals with reduced-form evidence.

Quantitatively, my results imply that optimizing rollover policy can lead to substantial welfare improvements for the principal. The precise magnitude of the gains depends on the welfare measure, but lies between 11 and 22% relative to a use-it-or-lose-it regime. These gains turn out to be slightly smaller than those predicted by a pure precautionary savings model for moving from use-it-or-lose-it to full rollover (Liebman and Mahoney, 2017).[3] Crucially however, my model indicates that the highest gains are not achieved by allowing full rollover, but by permitting agencies to retain about one third of their unspent funds at the end of the fiscal year. Thus, the model can explain the hesitancy of parliaments to allow full rollover, which is a very rare policy in practice.[4] Another reason to include a policy drift mechanism in the model is that, typically, agencies do not always spend their whole budget, which cannot be explained by a pure precautionary savings model. However, this budgetary slack mechanism turns out to have important welfare implications, as shown in Section 3.3.

The literature on the optimal timing of spending over budgetary cycles originates from studies on year-end spending spikes, such as those illustrated in Figure 1. This is not an isolated phenomenon occurring only for US federal contract obligations. Instead, it is so prevalent that it has become known under the nickname "December Fever" by people in the industry (Tarschys, 2003), even though the fiscal year of course does not end in December for all countries.[5] Eichenauer (2020) uses precisely this variation in fiscal timing to show that these spending spikes indeed seem to be caused by the budget cycle, rather than just being a product of an underlying spending seasonality. Most papers which document such a spending spike also find evidence that this type of expenditure rush at the end of the year tends to be highly inefficient. The most seminal paper in this respect is Liebman and Mahoney (2017), showing that US public procurement spending is on average 4.9 times higher in the last week of the fiscal year, but substantially more likely to be rated as low quality by an external evaluation. In addition, they find that competitively sourced government contracts at the end of the year are more likely to be awarded after receiving only a single bid, which lends further credibility to the narrative of rushed and less effective spending during the December fever. The latter point has also been made in the European context by McCue et al. (2021). Given the lower number of bids, it does not come as a surprise that there is also evidence that government agencies tend to pay a higher price for an otherwise comparable contract before the yearly budget expires (Klymak and Baumann, 2022). The

[3] In Liebman and Mahoney (2017), the compensating variation for optimized rollover relative to use-it-or-lose-it under CRRA utility is about 13.4% whereas in my model it turns out to be 11%.

[4] In fact, full rollover turns out to be just slightly less desirable for the principal relative to a use-it-or-lose-it regime in my analysis.

[5] In Canada, for example, the term "March Madness" has been coined instead.

combination of lower quality and higher volume of spending would be concerning to a principal aiming to optimally allocate tax payers' money, especially in the light of the large costs associated with inefficient procurement (see for example Best et al. 2023). Therefore, all the aforementioned papers suggest that a more even distribution of public spending throughout the year would be desirable.

Although the literature generally agrees on these *consequences* of year-end spending spikes, there is no such consensus when it comes to the *causes* which drive them. However, as will become evident in the theory section of this paper, the mechanism behind the spending spike is crucial for the decision whether agencies should be allowed to roll over their budget. The existing literature can be broadly categorized into two strands. This paper aims to understand the impact of rollover rules by combining both strands in a single model, which is essential for determining optimal rollover policy. One strand emphasizes *precaution*, while the other focuses on *procrastination*. The precautionary model is mostly represented by Liebman and Mahoney (2017). This model is characterized by uninsured risk faced by the agency in the form of productivity shocks and results in a precautionary savings model à la Aiyagari (1994). There are some crucial differences to a classic Aiyagari model[6] though, most importantly that income accrues only at the beginning of the fiscal year (i.e., every 12th period in a monthly model) and that there is no return to savings (as public bureaucracies generally cannot invest their budget). In this type of model, the spending spike occurs because agencies prefer to spend the precautionary savings accumulated throughout the fiscal year rather than letting them expire as unspent budget. In such a model, allowing the agency to roll over funds is unambiguously desirable, as shown by Liebman and Mahoney (2017). Intuitively, since the agent has the same flow utility as the principal, restricting rollover merely imposes a constraint on the principal's available savings technology. Adding constraints to the problem can only (weakly) decrease its optimal value, so allowing rollover is certain to be welfare-improving.

The literature following this seminal paper puts more emphasis on the possibility that the agency's incentives could be somewhat misaligned with the principal's, which in practice is the reason for which most budget authorities do not allow completely free rollover. To generate the year-end spending spikes, the misalignment is generally constructed such that spending is delayed within the fiscal year. Therefore, I bundle these models of policy drift under the umbrella term of *procrastination* models. The common feature among them is that they all assume that, for an exogenous reason, the agency's marginal utility of spending a given amount is in expectation higher at the end of the fiscal year than at the beginning. In Khalil et al. (2019) and Siemroth (2022) the agencies derive more utility from spending at the end of the fiscal year because it allows them to use the funds for self-serving rather than productive purposes. In Baumann (2019), the agency discounts the flow utility from production up to the end of the fiscal year, making early-year spending more heavily

[6]Sometimes also referred to as a Bewly, Hugett, or Imrohoroglu model.

discounted and thus of lower marginal value to the bureaucracy. Moreover, the agency faces an effort cost from spending large amounts in any given month, heightening the aforementioned discounting effect. In these models, constraining rollover is desirable since it imposes discipline on the agency's spending behavior, potentially aligning it with the principal's preferred spending policy, as is shown by Siemroth (2022) and Khalil et al. (2019). It is not immediately clear that this is also the case for the mechanism proposed by Baumann (2019), but I show in Proposition 2 that it has similar properties. However, these models are lacking the precaution channel, through which, for a given rollover rule, uneven spending may occur for reasons also desirable for the principal, i.e., optimally insuring against shocks.

The exception is Baumann (2019), which does include the precaution channel but is not concerned with rollover policy. Instead Baumann (2019) proposes a quasi-tax on the agency which increases over the fiscal year to an extent that incentivizes smoothed out spending.[7] However, in practice rollover is the much more active policy margin in public finance. One conceivable reason for this is that such a tax turns the effective annual expenses for the principal into a random variable, which would make budgeting vastly more difficult and costly.[8] Moreover, as will be highlighted by Proposition 1, perfectly even spending is also not optimal for the principal. And finally, the logic behind the tax incentive in Baumann (2019) relies on the assumption of i.i.d. productivity shocks. If shocks are auto-correlated, which my calibration suggests, designing a tax scheme to counter policy misalignment becomes vastly more difficult (as shocks are unobservable to the principal) and the issue of the agency's effective costs becoming volatile is heightened even more. Therefore, rollover remains the most prominent policy for parliaments to curb the within-year distribution of agency spending.

The remainder of this article is structured as follows: Section 2 reviews the existing models in the field and builds upon them to create a framework suited for the study of optimal rollover policy. Then, Section 3 calibrates the model, evaluates potential welfare gains and provides a reduced-form analysis of a rollover liberalization in Switzerland. Finally, Section 4 concludes.

[7] The budget spoken for the agency is increased in proportion to the expected tax revenue such that it does, on average, not affect the total amount of funds available for the agency.

[8] To some extent this is also true for partial rollover regimes in which the amount of returned funds to the principal is stochastic. However, this distribution is one-sided bounded as the effective costs may be lower than the spoken budget, but never higher.

2 A Model of Sub-Annual Spending Patterns

2.1 Motivation

To study the impact of rollover on agency behavior, I build on the precautionary savings model in Liebman and Mahoney (2017) in combination with a procrastination channel which drives a wedge between the principal's and the agency's flow utility. The primary reason to include such a wedge is to give the model additional flexibility to reflect the notion that restricting rollover might serve as a tool for the principal to align the agency's spending preferences with her own. After setting up the model in Section 2.2, I show in Proposition 2 that this channel becomes possible, but not mechanically enforced in my model. Moreover, as will become apparent throughout this paper, certain aspects of agency behavior (such as budgetary slack for instance) cannot be explained by a pure precautionary savings model.

To formalize this wedge, I decide to leverage the framework from Baumann (2019), which models procrastination by discounting the agency's utility from production to the end of the year and imposing an immediate disutility from procurement effort. As I will argue below, this approach of modeling procrastination bears close resemblance to the seminal dynamic model of procrastination in Akerlof (1991). Adding the possibility of rollover is one of the extensions to the model added in this paper since, as mentioned in Section 1, Baumann (2019) analyzes a tax rather than a rollover scheme. It should be noted that there are other approaches to modeling a procrastination channel in public finance, as highlighted in Section 1. Instead of explicitly modeling procrastination, one could let the agency exogenously gains more from spending at the end of the year (see Siemroth 2022 and Khalil et al. 2019). The primary reason for which I instead pick the formulation from Baumann (2019) is that I also aim to take into account the potentially systematic non-exhaustion of budget, a behavior which can only be generated by a model of procrastination which potentially breaks local non-satiation for the agency. The aforementioned literature has not modeled this explicitly, but in practice budgetary slack is an issue in the data to which I calibrate the model, and in public finance more generally. In fact, only about 80% of the consulting budget is used by the average agency in any given fiscal year. Budgetary slack is very common in multi-period budget setups and has been documented in many other contexts, both in the public and the private sector (Webb, 2002; Dunk and Nouri, 1998; Elmassri and Harris, 2011). This literature primarily evokes uncertainty as a reason for budgetary under-utilization, but from a modeling perspective there must be something breaking local non-satiation in order to produce a slack budget constraint for the agency, which cannot be achieved by elevated uncertainty alone. To the best of my knowledge, the only established approach to modeling procrastination in this literature which can satisfy this requirement is the one outlined in Baumann (2019). When evaluating welfare gains from rollover, budgetary

slack should be taken into consideration because the principal derives some value from returned funds and more even spending throughout the fiscal year may lead to the agencies either utilizing or returning more of their budget. In fact, as will become clear in Section 3, this is the main driver of welfare gains in the most favorable equilibrium that the principal can implement for my calibration.

Another reason for which I choose this approach to modeling procrastination is the close resemblance it bears to Akerlof (1991), which is arguably the most seminal dynamic model of procrastination. Similarly as in Baumann (2019), the return on an action depends on the distance from a deadline and there is an exogenous parameter which heightens present marginal effort relative to present marginal return. This introduces dynamic inconsistency into the model (on the side of the agency), which manifests as procrastination. In that sense, the way procrastination is modeled here is very similar to standard microeconomic approaches to procrastination.

2.2 Setup

Consider an agency operating on a budget $\mathcal{B}$, which is allocated over a time horizon $\mathcal{M}$ and re-instated every $\mathcal{M}$'th period. In each period t, the agency can produce a homogeneous output y_t subject to its production possibility frontier $y_t = z_t f(x_t)$, in which x_t is spending in period t and z_t is a productivity shock assumed to follow a log-normal Markov process with $\mathbb{E}[z_t] = 1$, variance σ_z, and a persistence parameter ρ_z. Even though $f(\cdot)$ is $\mathbb{R} \to \mathbb{R}$ here, the agency can be thought of as sourcing many factors of production and aggregating them in a cost-minimizing constant elasticity of substitution (CES) bundle. Thus, $z_t f(x_t)$ represents an optimal value function of a CES aggregation problem given contemporaneous spending x_t. More concretely, let there be a continuum of input varieties ω of measure one associated with quantities $q_t(\omega)$ and price $p_t(\omega)$. The agency's CES production function is $y_t = \max_{\{q_t(\omega)\}} \mathcal{A}_t \left(\int_0^1 a_t(\omega)^{\frac{1}{\epsilon}} q_t(\omega)^{\frac{\epsilon-1}{\epsilon}} d\omega \right)^{\alpha\left[\frac{\epsilon}{\epsilon-1}\right]}$ subject to the constraint $\int_0^1 p_t(\omega) q_t(\omega) d\omega = x_t$, in which $\mathcal{A}_t$ is total factor productivity, $a_t(\omega)$ is a potentially time-varying input-specific production weight, ϵ is the elasticity of substitution and α is a returns to scale parameter. Since this is a purely static problem, the agency solves this intra-temporal optimization separately from the inter-temporal problem of how to allocate spending, x_t, over time. It yields the following per-period production function:

$$y_t = \underbrace{\mathcal{A}_t P_t^{\alpha(\epsilon-1)} \left(\int_0^1 a_t(\omega) p_t(\omega)^{\frac{\epsilon(1-\epsilon)-1}{\epsilon}} d\omega \right)^{\alpha\frac{\epsilon}{\epsilon-1}}}_{z_t} \underbrace{x_t^{\alpha}}_{f(x_t)} \tag{1}$$

In which $P_t = \left(\int_0^1 a_t(\omega) p_t(\omega)^{1-\epsilon} d\omega \right)^{\frac{1}{1-\epsilon}}$ is the standard CES price index. One reason to microfound this production function is to better understand the drivers of the productivity shifter z_t. It could be affected

by total factor productivity or price level shocks, but also by correlated innovations to the input prices or production weights. Since all of those shocks would enter z_t multiplicatively, its overall variance could conceivably be substantially larger than the variance of macroeconomic aggregates such as the price level. I highlight this point because it relates to a reason for which Baumann (2019) discards the precautionary savings motive as a driver of spending spikes as in Figure 1. Namely, that it requires a variance of z_t that is orders of magnitudes larger than macroeconomic variables' volatility to explain year-end spending spikes, which he deems unrealistic. Equation 1 lends more credibility to a high σ_z as implied by the calibrations in Liebman and Mahoney (2017) and this study (see Section 3), as it is a product of several volatile factors and not only macroeconomic aggregates.

The agency also suffers a disutility from spending, as in Baumann (2019), denoted $e(x_t)$. It is additively separable and has CRRA form, similarly to the disutility from labor in a real business cycle model. It represents the significant effort and long hours associated with spending and procuring over capacity in a given month. Therefore, $e(\cdot)$ intuitively should be low for standard amounts of spending and rapidly increase above some subsistence level. One can think of this as a search cost for procurement similar to a standard search cost for labor in dynamic models featuring effort aversion (see e.g. Chetty 2006). The notion that procurement comes with substantial effort costs to be taken into account by a principal goes back to Tirole (1986) and Rogerson (2003) and is also empirically well-founded (Bajari and Tadelis, 2001).[9] One difference of these models as well as Khalil et al. (2019) to my setup is that they feature type uncertainty about the degree of the agency's effort aversion which can potentially be elucidated over a finite horizon. In my model, the problem is potentially infinite horizon, the principal is fully aware that the agency is effort averse, knows the shape of $e(\cdot)$, and takes it into account when allocating the budget. The only component of the agency's problem which is unobservable to the principal are the realizations of the shocks, z_t.

Turning to the definition of rollover, let $\mathcal{R}$ denote an agency's fiscal reserves and $k : \mathbb{R} \to \mathbb{R}$ be a rollover rule (i.e., a savings technology) at the end of the fiscal year, which maps unspent funds to additional budget in the next fiscal year. In addition, let $u(y_t)$ denote the agency's utility from production. If $k(\mathcal{R}) < 0$ for some $\mathcal{R} > 0$, this would correspond to the notion that leftover budget is "punished" by the principal by cutting the next fiscal periods funds upon observing leftover budget at the end of the year. Finally, β is the discount factor. Thus, we have defined all the elements necessary to write down the agency's value function given the state variables current budget B, fiscal period m and contemporaneous productivity z recursively:

[9]Public procurement in developed countries is highly regulated such that it can be extremely time- and effort consuming for the procuring agency. One example to highlight this, pointed out by Bandiera et al. (2009) is that US military procurement documents feature a 26 page description detailing the distinction between brownies and chocolate cookies.

$$v_A(B,m,z) = \begin{cases} \max_{x_t \in [0,B]} \beta^{\mathcal{M}-m} u(y(x,z)) - e(x) + \beta \mathbb{E}[v_A(B-x, m', z')|z], & \text{if } m < \mathcal{M} \\ \max_{x_t \in [0,B]} u(y(x,z)) - e(x) + \beta \mathbb{E}[v_A(1+k(B-x), 1, z')|z], & \text{if } m = \mathcal{M} \end{cases} \tag{2}$$

In this recursive representation, all variables with a prime superscript represent one period leads whereas variables without a prime are contemporaneous realizations. The agency's budget serves as the numéraire and is normalized to unity (this is without loss of generality, as will be elaborated when discussing the principal's problem below). Compared to a standard precautionary savings problem, a complication in the above Bellman equation is that the current fiscal period is an additional state variable in the system. The heightened discounting of $u(y(x,z))$ during the year comes from the formulation adopted from Baumann (2019), specifying that the agency discounts utility from public good production to the end of the year - when its production is observed and implicitly or explicitly evaluated by the principal. Relative to a standard precautionary savings model, this discounting structure leads to the agency postponing spending within the year, which is amplified by the disutility from effort aversion accruing immediately.

On the other side, there is a principal who cannot observe the productivity innovations z_t and transfers a budget to the agency every $\mathcal{M}$ periods. The principal is unaffected by effort aversion, but also values production. However, the principal does not discount production to the end of the fiscal year, but applies standard sequential discounting. Finally, the principal also considers her opportunity cost of budget, λ, when funds are allocated to the agency. Liebman and Mahoney (2017) calls this the "social value of budget". It is possible for the agency to end up with budgetary slack resulting in the return of some funds to the principal at the end of the fiscal year. These returned funds in period $\mathcal{M}$ reduce the disutility from increasing the budget by λ since it effectively lowers the budget required by the agency for the next year. Note that, since $\mathcal{R}$ are the reserve funds leftover at the end of the fiscal year, and $k(\mathcal{R})$ is the rolled over budget, $\mathcal{R} - k(\mathcal{R})$ gets returned to the principal. During the fiscal year, the principal must rely on the agent to produce, but at the end of the fiscal year she decides on the next year's budget for the agency, denoted $\mathcal{B}$. In its recursive form, the principal's value function is:

$$v_P(B,m,z) = \begin{cases} u(y(x,z)) + \beta \mathbb{E}[v_P(B-x, m', z')|z], & \text{if } m < \mathcal{M} \\ \max_{\mathcal{B}>0} u(y(x,z)) - \lambda(\mathcal{B} - [\mathcal{R} - k(\mathcal{R})]) + \beta \mathbb{E}[v_P(\mathcal{B} + \mathbb{E}[k(\mathcal{R})], 1, z')|z], & \text{if } m = \mathcal{M} \end{cases} \tag{3}$$

Instead of normalizing $\lambda = 1$, like in Liebman and Mahoney (2017), I normalize $\{\lambda \in \mathbb{R} | \mathcal{B} = 1\}$, i.e., I define λ to take on a value which leads the principal to set the budget to unity. This is without loss of generality, but simplifies the numerical simulation later on. More importantly, as I show in Appendix

5.1, normalizing $\mathcal{B}$ pins down a unique λ, but normalizing λ may not pin down a unique $\mathcal{B}$. Thus, it is conceptually more appealing to use $\mathcal{B}$ as the numéraire in my setup. This issue is related to the possibility of budget being returned to the principal, thus it works in Liebman and Mahoney (2017) to normalize λ because in their model, the agency always spends all available budget in equilibrium. The result is summarized in the following lemma:

Lemma 1 *Assume that effort aversion and flow utility are continuously differentiable, $\alpha \leq 1$ with $u(\cdot)$ concave and $e(\cdot)$ convex. Then, there always exists a unique $\lambda \in \mathbb{R}_{++}$ such that the principal finds it optimal to assign a unity budget to the agency. For a given λ however, the implicitly defined optimal budget may not be unique.*

The proof can be found in Appendix 5.1. Equation 3 will be crucial to value counterfactual spending sequences by the agency from the perspective of the principal. Notice that if $m = \mathcal{M}$ and the principal decides on a budget, it is the long-run expectation of rolled over budget entering the next period's value function rather than the realized $\mathcal{R}$. The reason for this is that it is not possible for the principal to condition the budget on the realized value of rollover, $k(\mathcal{R})$, and still have the agency implement a different spending sequence than it would under a use-it-or-lose-it regime. Therefore, the second best option for the principal is to make the budget dependent on the long-run expectation of rolled over funds if she wants to deviate from the no-rollover spending equilibrium. This is essentially an application of Proposition 4 in Liebman and Mahoney (2017). The intuition behind this result is that the agency anticipates that the principal is going to adjust $\mathcal{B}$ to allocate its desired level of funding, $\mathcal{B}^*$, to the agency depending on the rolled over amount. Therefore, the agency would start the next fiscal year with $\mathcal{B}^*$, no matter the level of reserves at the end of the year. Anticipating that, the best response for the agency is to deviate to its spending policy under no-rollover, i.e., operating under the assumption that leftover funds will not affect its budget constraint in the following fiscal year.

A first and crucial insight which this model delivers concerns the sub-annual timing of spending in the face of uncertainty. Baumann (2019) highlights in Proposition 2 that equal spending in fiscal periods throughout the year is optimal for the principal in a deterministic two-period model. Consequently, the premise that observing heightened spending towards the end of the fiscal year must be inefficient for the principal has prevailed. However, I highlight in Proposition 1 that this is no longer true in the face of uncertainty.

Proposition 1 *If the principal were able to implement the spending sequence given the allocated budget and perfect rollover, expenditure would on average not be even throughout the fiscal year as long as $\sigma_z > 0$ and $u(y(\cdot))$ is concave.*

Therefore, simply observing an uneven spending pattern throughout the year as it is displayed in Figure 1 does not necessarily contradict optimality for the principal. The proof for Proposition 1 can be found in Appendix 5.2. This is not to say that the pattern highlighted in Figure 1 must be efficient, in fact the literature reviewed in the introduction suggests substantial inefficiencies, but it cannot be concluded that an on average more even spending sequence is necessarily more desirable for the principal. In practice, I show in Appendix 5.2 that in the context to which I calibrate the model, the principal would indeed prefer somewhat higher spending at the end of the fiscal year. However, this heightened optimal spending is not as sudden or pronounced as the spike we observe implemented by the agencies in the US data from Figure 1 or the Swiss data to which the model is calibrated.

The intuition behind Proposition 1 is simple: If the period in which the budget gets reinstated lies far in the future, there is a relatively high probability to experience a sequence of adverse shocks before the new budget arrives. During the late fiscal year, however, the scenario of a series of low productivity shocks is less daunting, since new budget is about to arrive which can be used up in these periods to avoid hitting the liquidity constraint, even if most currently available funds are spent. Therefore, the precautionary savings motive is more pronounced in the early periods of the fiscal year, even if there is perfect rollover, no effort aversion, and sequential discounting as in Equation 3. This proposition highlights also that just because spending increases towards the end of the year, it cannot be concluded that there must be misaligned incentives between the agency and the principal, since such a pattern is in fact a completely natural property of optimal spending behavior in a precautionary savings model with multi-period budgeting.

In practice however, the principal is still often concerned about there being a wedge between their own and the agency's flow utility, which is why parliaments usually are reluctant to grant agencies full rollover rights. A principal may prefer increased spending towards the end of the year, but not of the magnitude observed in the data (see e.g., Figure 1). The prevailing argument is that under restricted rollover, the more direct control the parliament has over the agency's budget constraint leads to spending decisions which are closer to the principal's preferred policy (see for example Proeller 2021). If this is of concern to practitioners, the model should reflect the possibility of this channel being at work when the agency's preferences differ strongly from the principal's. Proposition 2 shows that this can be the case in the model constructed here, but only if the wedge (marginal effort aversion) is pronounced enough. In other words, adding effort aversion to the model is sufficient to enrich it with the possibility of an "alignment" effect of rollover restrictions.

Proposition 2 *Starting from a full rollover equilibrium, the subset of the state space during fiscal periods $m < \mathcal{M}$ for which constraining rollover will weakly decrease the distance between the agency's and the principal's optimal spending policy, $|x_A^* - x_p^*|$, is increasing in $e'(\cdot)$.*

Proposition 2 highlights why in this model, the principal may find it optimal to constrain the agency's year-end savings technology. The proof can be found in Appendix 5.3. Intuitively, if the wedge is sufficiently pronounced, there will be more situations in which constraining rollover aligns the agency's spending preferences with the principal's. There is a trade-off between allowing the agency to save optimally to build up a buffer-stock in response to shocks and disciplining their spending policy. In Liebman and Mahoney (2017), only the first channel is present, such that unconstrained rollover is always the preferred policy by the principal. Another refinement relative to Liebman and Mahoney (2017), which potentially mitigates the gains from full rollover, is that my model allows for the possibility that some budget is returned to the agency, which is also valued at the social cost of funds. As highlighted in Section 2.1, budgetary slack is very common in public sector bureaucracies, and also the data to which I calibrate the model features average budget under-utilization of about 20%. When calibrating the model in Section 3, I also match this moment, such that less generous rollover rules might generate value for the principal by having the funds returned rather than the agency spending it at a marginal value lower than the social cost of funds, i.e. the principal's opportunity cost of budget. To evaluate the relative weight of these channels, the next section will calibrate the model to a use-it-or-lose-it context and perform counterfactual simulations of alternative rollover policies.

3 Empirical Analysis

3.1 Data & Institutional Context

In order to apply the model developed in Section 2 consistently, the institutional context must correspond reasonably well to the setup of the model. Most importantly, the focus should lie on a spending sequence for which the agency has discretion over the timing of expenses. For example, it is undesirable to have permanent employees' salaries or infrastructure servicing bills included in the target sequence, since spending is likely to follow (at least to a large extent) a predetermined pattern over which the agency does not have direct control. This is one of the reasons why Liebman and Mahoney (2017) focus on government contracts for external suppliers. In addition, particularly in my model in which budgetary slack can occur, it is important that the analyzed spending sequence is at the same level as the budget constraint binds. I.e. when analyzing procurement spending, the relevant budget allocated to the agency should also be specific to that type of procurement.[10] The reason for this is that picking a spending series at a lower level than the budget is spoken might not capture the relevant optimization problem, since then also substitution patterns between

[10] Baumann (2019) and Liebman and Mahoney (2017) implicitly assume that the observed total spending within a fiscal year is equal to the agency's budget.

varieties $q(\omega)$ in the observed series and other varieties for which the budget is used influence the expenditure decision. Intuitively, one could not be sure if a spending adjustment comes from the intra-temporal problem in Equation 1 or from the inter-temporal one in Equation 2. However, this is relevant for evaluating the optimality of this adjustment from the principal's perspective. For this reason, the institutional context for most Swiss public bureaucracies prior to 2017 is a favorable domain of application. Federal agencies were subject to a strict no-rollover regime and had specific sub-budgets allocated to them by the parliament on an annual basis for certain types of procurement. The two types of procurement for which a wide range of agencies have a specifically allocated budget and satisfy the above requirements are external consulting and IT contracts. Unfortunately, the sub-annual data for the latter is systematically incomplete,[11] so I focus my analysis on consulting spending, which encompasses about 112 million USD per year in my data and is spread across 41 federal agencies.[12]

A further advantage of this institutional context is that there was a budgeting reform introduced in a staggered fashion between the early 2000s and 2017. This reform titled "New Leadership Model for the Federal Administration"[13] introduced the possibility for agencies to roll over some funds and build up a buffer-stock of fiscal reserves. Due to its staggered roll-out and permanence once in place, this reform allows me to credibly support the counterfactual analysis with reduced form impact evaluations, validating its consistency with the model's predictions. Though the reform allowed only for relatively constrained rollover, I find that the reduced form evidence confirms the model's prediction for sub-annual spending patterns after a partial rollover liberalization. Details on this analysis as well as the reform itself can be found in Section 3.4.

The data on budgets of federal bureaucratic agencies is publicly available.[14] Spending data on the agency level is only public on the annual frequency, but monthly data can be obtained from the federal department of finance upon request. Data on the amount of reserves built by the agencies after the 2017 reform was collected from the federal state accounts, which also are publicly available.[15]

[11] Incomplete in the sense that the spending data registered at the Swiss Department of Finance does not encompass all IT procurement spending, such that the sub-annual spending data does not necessarily sum up to the total annual spending. If the probability of spending being non-registered is not uniform throughout the fiscal year, this could invalidate my analysis. Therefore, I restrict myself to consulting spending, for which the sub-annual data consistently sums up to the annual total expenses.

[12] Note that this is only about half of the overall consulting expenses by federal agencies since some types of specialized consulting do not accrue to the general consulting budget. For example, certain kinds IT consulting are booked under the IT budget such that they do not show up in my data.

[13] The German original: "Neues Führungsmodell für die Bundesverwaltung"

[14] Check here for budgets dating back to 2003:
https://www.efv.admin.ch/efv/de/home/finanzberichterstattung/finanzberichte/budget.html

[15] Consider Section 2A and 2B in the state accounts dating back to 2017 published here:
https://www.efv.admin.ch/efv/de/home/finanzberichterstattung/finanzberichte/staatsrechnung.html

3.2 Calibration

I calibrate the model to the time period in which agencies were subject to a no-rollover regime, since this is the most clear-cut case implying that $k(\mathcal{R}) = 0$ for all $\mathcal{R}$. In addition, there is the most data available for the use-it-or-lose-it regime as rollover was only liberalized in 2017. This setup of the calibration involves some implicit assumptions to be discussed below before proceeding to the parameter estimation.

Ratchet Effect. One might argue that if there is no rollover *de jure*, there could be *de facto* negative returns on $\mathcal{R}$ in the sense that upon observing unspent funds leftover at the end of the fiscal year, the principal would reduce the agency's budget going forward. In that case, one would need to set $k(\mathcal{R}) < 0$ for the baseline calibration. This can be seen as a type of Ratchet effect (Freixas et al., 1985), since the agency's savings potentially send a signal about its productivity to the principal. However, Liebman and Mahoney (2017) provide some reasoning why this channel is absent in their model, as it may not be optimal for the parliament to condition allocated funds on it if budget setting is a repeated game,[16] which also applies in this extended model. In addition, the parliament's behavior in the data does not seem to exhibit this kind of Ratchet effect, as there is no statistically significant relationship between the share of leftover funds at the end of the fiscal year and the following year's budget allocated to the agency (the corresponding analysis is documented in Appendix 5.4). Therefore, I impose the $k(\mathcal{R}) = 0 \; \forall \, \mathcal{R}$ rule for the baseline calibration.

Limited Commitment. As previously discussed in Section 2, the principal can only implement an equilibrium spending sequence different from no-rollover by committing to not conditioning the following year's budget on a given year's realization of $\mathcal{R}$. One implicit assumption is that the principal can make this commitment with full credibility. If this commitment were only partial, i.e., the principal would deviate from it with some probability π, this would also need to be taken into account in the calibration, as it is well-knows to impact the achievable welfare gains in principal-agent models (Sappington, 1991; Rey and Salanie, 1996). In the context of this paper, the principal could simply reduce the agency's budget proportional to the amount of funds rolled over if she were unable to commit.[17] Therefore, I collect data on agency's budgets and rolled over reserves following the 2017 reform and test for lacking parliamentary commitment. If that were the case, one would expect to observe a negative correlation between the amount of funds rolled over and the growth rate of agencies' budgets. However, I do not find evidence for such an association across a range of specifications, as documented in Appendix 5.4. Given this evidence, I proceed under the premise that $\pi = 1$, i.e., Parliament has the ability to fully commit to a budget setting rule.

[16] See Footnote 44 in Liebman and Mahoney (2017).

[17] There is also some discussion of this issue in Liebman and Mahoney (2017), though they call it "raiding" the rolled over funds. In the model, reducing the following budget and raiding rollover funds is equivalent. In the data however, we would expect to observe the first since the principal has freedom to set the agency's budget, but is legally unable to take control of its reserves.

In terms of parameterization, I take $\mathcal{M} = 12$ since my data has monthly frequency and therefore impose $\beta = 0.996$ to get a standard value for annual discounting of around 5%. Thereafter, following Baumann (2019) I assume that $u(\cdot)$ takes the simplest possible CRRA form, such that $u(x) = x^{\phi}$ in which ϕ controls the coefficient of relative risk aversion as well as the inter-temporal elasticity of substitution.[18] Given the production function derived in Section 2, the flow utility from production thus becomes $z_t x^{\phi\alpha}$. The calibration cannot identify α and ϕ separately, but only their product, i.e, the exponent on flow utility from production, denoted $\varphi = \alpha\phi$.[19] For the shock z_t, both its variance σ_z and its persistence parameter ρ_z need to be calibrated. Effort aversion is also parameterized as simply as possible, letting $e(x) = \Gamma x^{\eta}$ for some $\eta \geq 1$ and $\Gamma \geq 0$. This follows the standard way to model disutility from labor or effort with Γ governing its weight relative to utility from production and η controlling its convexity.[20] This leaves the vector $\Theta = \{\varphi, \sigma_z, \rho_z, \eta, \Gamma\}$ to be calibrated.

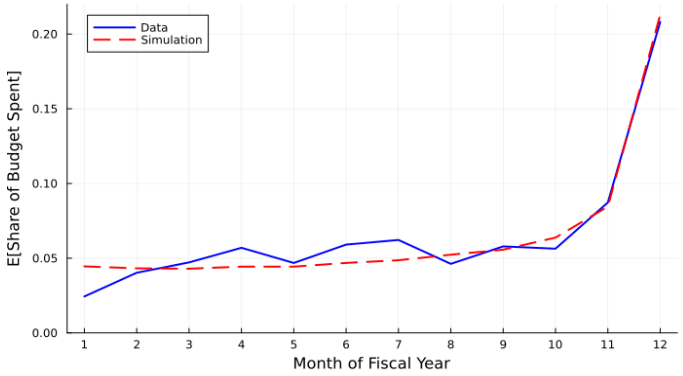


Figure 2: Expected share of available consulting funds spent in each fiscal month in the data and the calibrated model.

To this end, I resort to a simulated method of moments (SMM) calibration, targeting the expected share of available funds spent at the end of each trimester as well as the first month of the fiscal year (i.e., the first, fourth, eighth and twelfth fiscal period). In addition, I target budgetary slack as the average share of funds returned to the parliament at the end of the year, leaving me with five moments to match with the five parameters in Θ. The federal spending data is available from 2006 onward, but only periods up to 2017 are used for the calibration since subsequently rollover was liberalized universally. I observe consulting budgets for 42 agencies, but only keep observations for which the sequence covers the full fiscal year.[21] Also

[18] Computationally, this is somewhat more convenient than the standard CRRA formulation, since x is a share of annual budget and therefore quite small in some months. As the volatility of a standard CRRA function close to zero is very large, much longer simulations would be required to produce accurate counterfactuals. This choice is however inconsequential for the substantive results concerning welfare.

[19] Yet, I differentiate between the two parameters on a theoretical level to emphasize that concavity of the flow utility could be a consequence either of decreasing returns to scale in production or of concave Bernoulli utility from production.

[20] Baumann (2019) imposes linear effort aversion, stating that this does not restrict the model since only its relative magnitude of η to φ matters. However, partly because I allow for budgetary slack, this is not the case in my model, so φ and η have to be calibrated separately.

[21] If a budget was for example allocated due to a specific shock and only took effect in the middle of the fiscal year, these six

the agencies for which the budgeting reform takes effect before 2017 are excluded from the year in which they are allowed to roll over onward. This leaves me with 285 agency-year and 3420 agency-month pairs to use for the SMM. Since in the no-rollover case, each year is an independent optimization problem from the agency's perspective, I use backward induction to solve the model for each parameter guess and update guesses using the Nelder-Mead optimization algorithm, minimizing the sum of squared errors between the model's simulation output and the moments from the data.[22]

Table 1 reports the resulting estimates jointly with the fit on the targeted moments and a set of non-targeted second moments. Figure 2 visualizes the resulting expected spending in each fiscal period in the model simulation and the data. During fiscal periods one through nine, the observed data somewhat fluctuates around the essentially constant expected spending implied by the model simulation. I do not consider it a major issue that these fluctuations are not precisely replicated by the model, as that would arguably overfit the prediction to random perturbations. The heightened expected spending during the last month of the fiscal year can also be matched nicely. Moreover, the model replicates the elevated levels of spending during fiscal months ten and eleven, though this rise was not explicitly targeted. Also the (non-targeted) increase in spending volatility throughout the fiscal year as well as its level are reproduced remarkably well by the model, as displayed in the non-targeted moments section of Table 1.[23] I chose to report these second moments as non-targeted validation checks since their importance in a stochastic no-rollover model was underlined by Liebman and Mahoney (2017).

In terms of the estimated parameters, the calibration results appear to be quite reasonable and none of the parameters approaches their admissible boundary. In their calibration, Liebman and Mahoney (2017) only have two parameters to estimate, namely the curvature of the flow utility and volatility of the shock. The first is not directly comparable between their and my model, since here the curvature of the agency's flow utility also depends on effort aversion and varies by month of the year. The estimated σ_z is remarkably similar however, 1.7 in my model and 1.73 in Liebman and Mahoney (2017). Perhaps surprisingly, the persistence of the shock is implied to be quite low (about 0.34). However, even this moderate ρ_z has a major impact on precaution, as my data features a larger spending spike than the contracts in Liebman and Mahoney (2017), but is matched with almost the same shock volatility.

months would be excluded from the calibration.

[22] Of course, the backward induction approach only works while there is no rollover such that the optimization horizon is finite. For the counterfactual simulations allowing some rollover in Section 3.3, I solve the model relying on standard value function iteration for each equilibrium.

[23] I report medians rather than means since both the simulation and the data feature some large outliers such that the means of sub-annual volatility are very unstable.

Parameter	**Value**	**Method**
φ	0.453	Calibrated
η	1.105	Calibrated
Γ	1.367	Calibrated
σ_z	1.7	Calibrated
ρ_z	0.337	Calibrated
β	0.996	Imposed
$\mathcal{M}$	12	Imposed
Targeted Moments	**Simulation**	**Data**
Expected spending $m = 1$	0.044	0.024
Expected spending $m = 4$	0.044	0.056
Expected spending $m = 8$	0.052	0.046
Expected spending $m = 12$	0.212	0.208
Expected returned budget	0.215	0.207
Non-Targeted Moments	**Simulation**	**Data**
Median Volatility Ratio	3.035	2.819
Median Q1 Spending Volatility	0.016	0.017
Median Q2 Spending Volatility	0.015	0.017
Median Q3 Spending Volatility	0.023	0.022
Median Q4 Spending Volatility	0.097	0.083

Table 1: SMM calibrated parameters with targeted & non-targeted moments. The simulation spans 50'000 fiscal years. The first 1000 years are discarded as a burn-in period. The volatility ratio refers to the standard deviation of spending during the second semester of the fiscal year relative to the first.

3.3 Counterfactual Results

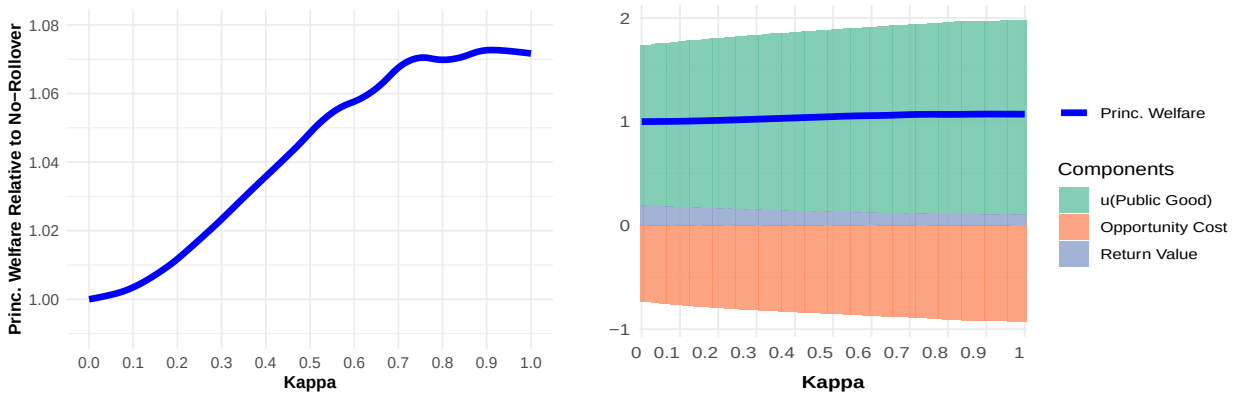
Given this calibration, I proceed to solve the principal's and the agent's problem for alternative rollover rules to explore potential welfare gains for the principal in these counterfactual equilibria.[24] Concretely, I consider two ways, in which rollover could be liberalized. First, one could allow the agency to roll over any amount up to some fraction of its allocated budget. In that case, the savings rule is discontinuous as funds can freely be carried over until the upper bound for reserves is reached, after which everything is returned to the principal. Such a policy is for example in place in Canada, where agencies are permitted to roll over up to 5% of their budget. Denoting κ as a parameter governing the fraction of the allocated budget which agencies can carry over, this implies that $k(\mathcal{R}) = \min\{\mathcal{R}, \kappa\mathcal{B}\}$. An alternative and even simpler policy would be that agencies can roll over a fraction of their amassed reserves, such that $k(\mathcal{R}) = \kappa\mathcal{R}$. To compute the value of these equilibria to the principal, her opportunity cost of allocating the budget, λ, is an important quantity. Given that the annual budget serves as the numéraire, the principal's first order condition for $\mathcal{B}$ implies that:

[24] Allowing for some rollover turns the optimization from a finite to an infinite horizon problem. Therefore, these equilibria are computed using standard value function iteration as opposed to the calibration, which used backward induction.

$$\lambda = \beta \mathbb{E}\left[\frac{\partial}{\partial \mathcal{B}} v_P(1 + \mathbb{E}[k(\mathcal{R})], 1, z')|z\right]\left(1 + \frac{\partial \mathbb{E}[k(\mathcal{R})]}{\partial \mathcal{B}}\bigg|_{\mathcal{B}=1}\right) \tag{4}$$

This formulation assumes that when the principal audits the agent at the end of the year and decides on a budget, she also observes z_M (though she does not observe the other shocks throughout the fiscal year) and therefore forms an expectation conditional on the shock in period $\mathcal{M}$. This allows the budget (or λ) to fluctuate from year to year for a given rollover rule, which is also the case in the data. However, if the principal is not allowed to observe $z_{\mathcal{M}}$, such that the budget is instead set based on the unconditional expectation, the results on welfare, relative to zero rollover, reported below remain practically unchanged. Intuitively, λ can be thought of as the principal's opportunity cost of the expended budget. It can change under a liberalized rollover policy because the expected amount of returned funds, the optimal budget, or its marginal value evaluated at $1 + \mathbb{E}[k(\mathcal{R})]$ changes. Since it can occur that some budget is returned to the principal, she also takes the effect of the expected returned funds into account when making the budget decision.[25] Consistently with Equation 3, the λ resulting from Equation 3.3 is also used to value the budget returned to the principal.

The results for the principal's welfare relative to the no-rollover scenario are displayed in Figure 3 for



(a) Principal Welfare

(b) Principal Welfare & Decomposition

Figure 3: Counterfactual results for $k(\mathcal{R}) = \min\{\mathcal{R}, \kappa\mathcal{B}\}$. Welfare is measured using the Lucas consumption equivalent factor as defined in Lucas (1987). The decomposition of the principal's value into the components on the right is based on Equations 3 and 2.

the budget fraction rule ($k(\mathcal{R}) = \min\{\mathcal{R}, \kappa\mathcal{B}\}$) and in Figure 4 for the savings fraction rule ($k(\mathcal{R}) = \kappa\mathcal{R}$) respectively. The panels on the left-hand side show the principal's welfare relative to the baseline no-rollover scenario, measured in the sense of a Lucas consumption equivalent (see Lucas 1987, 2003). Thus, the welfare

[25] Though she cannot make the budget contingent on the realized returned funds in a given year, as discussed in Section 2.2.

measure can be interpreted as the factor by which public good investment under no rollover would have to be scaled up in each period to make the principal indifferent to the equilibrium under an alternative rollover rule. Even though the Lucas consumption equivalent is often used in dynamic models, one could consider alternative welfare measures such as dynamic variants of the compensating and the equivalent variation (Liebman and Mahoney (2017) report a dynamic compensating variation). Therefore, I also present the welfare results displayed in Figures 3a and 4a in terms of these alternative measures in Appendix 5.5. While the scale varies somewhat between the different welfare measures,[26] the direction and qualitative patterns are very similar across them.

The right-hand side panels of Figures 3 and 4 displays welfare jointly with a decomposition into the value stemming directly from public good production, budget returned by the agency, and disutility incurred from the budget's opportunity cost. For the no-rollover equilibrium (i.e., $\kappa = 0$), these components sum to one and scale up or down proportionally to the component's contribution to the principal's value in the other equilibria. Evidently, the two policies have very different effects, though welfare gains are possible with either of them.

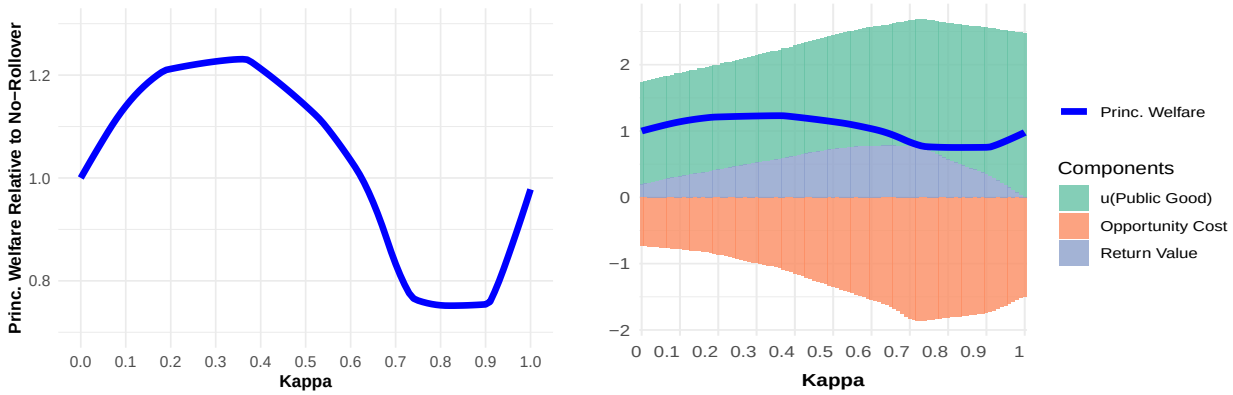
Allowing the agency to roll over a fraction of the allocated budget results in, perhaps surprisingly, little change in the principal's welfare, which is illustrated in Figure 3. As reflected by the consistently low value from returned budget, such a policy simply does not lead to the agency accumulating a very high level of reserves. There is a slow but steady increase of the utility from public good production as κ rises. The value of returned budget somewhat diminishes, but constitutes a negligible part of overall welfare. Finally, the opportunity cost also slowly increases, but at a somewhat lesser pace than the utility from public good production. Overall, this implies that allowing the agency to roll over up to an amount of almost their annual budget is optimal for the principal. More specifically, welfare gains of about 7% can be achieved by allowing the agency to roll over up to 92% of its annual budget. However, the gains from rollover become almost flat for κ larger than 0.7. The simulated agency behavior underlying the calculations in Figure 3 displays somewhat smoothed out spending throughout the fiscal year for the optimal $\kappa = 0.92$ equilibrium relative to no-rollover.[27] Moreover, the volatility of spending is more constant throughout the fiscal year in the $\kappa = 0.92$ equilibrium, though the overall standard deviation of spending remains virtually unchanged. The relative volatility of second to first semester spending drops from about 3 to 2.14, which corresponds to an a reduction of about 29%. This is to say that, even though introducing a $k(\mathcal{R}) = min\{\mathcal{R}, \kappa\mathcal{B}\}$ rule has relatively little impact on the principal's welfare, the agency's sub-annual spending patterns still change

[26] Most notably, the gains for the $k(\mathcal{R}) = \kappa\mathcal{R}$ rule are only about half as large for the equivalent and compensating variation as in terms of the Lucas number.

[27] Though $\mathcal{M} = 12$ spending is still more than twice as high as during the rest of the fiscal year, this is an almost 50% reduction of the spending spike relative to the no-rollover equilibrium.

quite substantially.

Overall however, allowing agencies to roll over a fraction of their savings rather than up to a fraction of their annual budget, appears more desirable for the principal. As displayed in Figure 4, such a policy has a much more pronounced impact on welfare and can lead to gains of up to 23% (in the sense of the Lucas consumption equivalent). The optimal κ which implements this equilibrium is 0.34, i.e., allowing agencies to roll over approximately one third of the funds available to them at the end of the fiscal year. Numerically, 23% sounds like a very high number, but when quantifying welfare with the compensating variation[28] rather than the Lucas number, it drops to about 11.2%, which is relatively close to the full rollover gains of 13% in Liebman and Mahoney (2017). Interestingly, the composition of this welfare gain is quite different from that of the optimized $k(\mathcal{R}) = \min\{\mathcal{R}, \kappa\mathcal{B}\}$ policy displayed in Figure 3. For the $k(\mathcal{R}) = \kappa\mathcal{R}$ policy, the gains at the optimum do not stem as much from increased utility due to public good production as from the returned budget. In other words, optimizing rollover in this case does not lead to much more or smoother public good provision. Instead, gains are made by using the available budget more efficiently, returning it rather than spending it wastefully and still producing the same as under no-rollover. This is reflected in the steep increase of the value from returned funds for lower values of κ in Figure 4, whereas the utility from public good production remains almost constant over this domain. For very liberal rollover, the value of returned budget must of course drop drastically as by definition zero budget is returned if κ equals unity. The direct utility from public good provision remains close to constant and then rises rapidly for $\kappa = 0.75$



(a) Principal Welfare

(b) Principal Welfare & Decomposition

Figure 4: Counterfactual results for $k(\mathcal{R}) = \kappa\mathcal{R}$. Welfare is measured using the Lucas consumption equivalent factor as defined in Lucas (1987). The decomposition of the principal's value into the components on the right is based on Equations 3 and 2.

[28] Similarly to Liebman and Mahoney (2017), the compensating variation in this context refers to the change in the agency's annual budget which renders the principal indifferent between an equilibrium and the no-rollover baseline, while keeping the principal's expenses constant.

and higher. However, this upsurge is not strong enough to compensate for the reduced returned budget and increased opportunity cost in these equilibria. For completely free rollover, the opportunity cost is somewhat reduced, which results in an uptick of the principal's value towards $\kappa = 1$. Intuitively, the opportunity cost captures how much budget the principal needs to allocate to make the agency implement the best achievable equilibrium under a given rollover rule. As highlighted by Equation 3.3, the opportunity cost can shift due to a number of factors in equilibrium. However, it seems that its downturn as $\kappa \to 1$ in Figure 4 is primarily driven by a strong increase of the expected amount of reserves rolled over as κ approaches unity. In principle, this could either increase or decrease λ as it implies that $\frac{\partial \mathbb{E}[k(\mathcal{R})]}{\partial \mathcal{B}}|_{\mathcal{B}=1}$ increases, but the first term decreases since the value function is concave in its first argument and its derivative gets evaluated at a higher point. It turns out that the latter effect dominates λ for very liberal rollover in my calibration. However, this is not the point the domain of κ in which the principal would like to locate in the first place, since full rollover is still slightly less desirable than use-it-or-lose-it and just imposing mild savings constraints can result in large welfare losses. Thus, one difference of this policy to the one displayed in Figure 3 is that erring in κ can result in substantial welfare losses, of up to almost 20%. When optimizing κ, also the agency spending behavior changes quite substantially, even more so than for the optimized fraction-of-budget rollover rule. Relative to the baseline no-rollover rule, the median ratio of second to first semester spending volatility drops by 42% to 1.73. The impact on the year-end spending spike is similar as for the optimized fraction-of-budget rule, i.e., a reduction of about 50%, such that spending in the last month of the fiscal year remains about twice as high as during the average earlier period. Differently than the fraction-of-budget policy however, the optimized $k(\mathcal{R}) = \kappa\mathcal{R}$ rule reduces also the median spending volatility quite substantially, by about 25% to 0.106. It seems that the discontinuity in the fraction-of-budget rule prevents the agency from adapting its behavior as strongly as in the fraction-of-reserves rule.

In summary, the optimal policy among the set analyzed in this section is to allow the agency to roll over about one third of its available funds at the end of the year, giving rise to welfare gains of around 23% for the principal. At the optimum, gains are primarily driven by returned budget to the principal, while utility from public good production is only heightened slightly. These results explain why in practice, parliaments are hesitant to grant very liberal rollover rights to government agencies, as setting κ too high can lead to large welfare losses. At the same time I show that by completely banning rollover, one misses out on quite substantial welfare gains from more efficiently spent and returned budget.

Though the degree to which these findings are generalizable to other contexts is unclear, the methodology can be implemented to pin down optimal rollover policy wherever sub-annual data is available and the institutional setup matches the model. However, the optimal discretion for rollover in Swiss public bureaucracies is likely to be close to the upper bound of optimal rollover discretion since, in general, higher legal discretion

is more effective for public procurement if institutions are of good quality (Bosio et al., 2022). Thus, one would expect optimal rollover policy to lie somewhat below one third of the accumulated reserves in contexts featuring lower-quality institutions.

3.4 Reduced-Form Evidence on Spending Spikes

This section goes on to analyze a policy liberalizing rollover for Swiss federal agencies and relates it to the results obtained above. As mentioned in Section 3.1, up until 2017 most Swiss public bureaucracies essentially operated under a "use-it-or-lose-it" rule, returning all unspent budget to the parliament. In 2017, a new system was put in place which allows to roll over funds subject to some justification to the parliament, but without an explicit upper bound on the amount which agencies can save. A direct before-after comparison of spending patterns would risk to be confounded by time-varying unobservables, but fortunately, 20 "pilot" agencies were already exposed to the reform earlier at various points during the preceding 19 years.[29] This staggered introduction provides a relatively favorable setting to test some of the predictions from the structural model in the data, albeit subject to some caveats. First, the agencies for which the policy was introduced earlier were most likely not randomly selected. However, they should also not be completely unrepresentative of the total population of agencies, because the reason for the staggered introduction was to get an impression of the policy's effect once applied to the entire federal administration. In addition, the following empirical analysis employs a staggered Difference-in-Differences design estimator (Sun and Abraham, 2021), such that the main assumption for consistency is not perfect randomization but conditional parallel trends.

Another concern is that in the same reform, the level of budget allocation was shifted to the agency, such that there were no longer purpose-specific sub-budgets within any given bureaucracy. This would be problematic if the agencies were to change the areas to which they allocate their budget in response to the reform, potentially confounding the effect of liberalized rollover. This concern remains, but can be at least somewhat alleviated by the fact that I do not find evidence for a systematic change in the relative spending on different purposes in response to the reform (see Appendix 5.6).

Though the reform liberalized rollover, the amount of funds which agencies can retain is not clearly defined, but decided on a case-by-case basis. Therefore, the new equilibrium does not directly map into one of the cases analyzed in Section 3.3. Yet, verifying that the response of agencies' spending patterns goes into the direction which the model predicts for partially liberalized rollover provides a valuable coherence check. Baumann (2019) argues against the relevance of the precautionary savings channel by showing that in his sample (government agencies in the UK), more liberal rollover regimes were not accompanied by reductions

[29] More details (in German) as well as a qualitative evaluation of the reform can be found in Proeller (2021).

in spending spikes. If that were the case, the reduced from evidence would stand in strong contrast to the predictions of my model (as well as Liebman and Mahoney 2017), since the simulations in Section 3.3 feature more even spending relative to the use-it-or-lose-it pattern displayed in Figure 2 for more free rollover regimes. In my analysis however, spending spikes in fact do seem to decrease in response to the reform, as displayed in Figure 5. Though spending spikes are measured with substantial noise, in the medium run

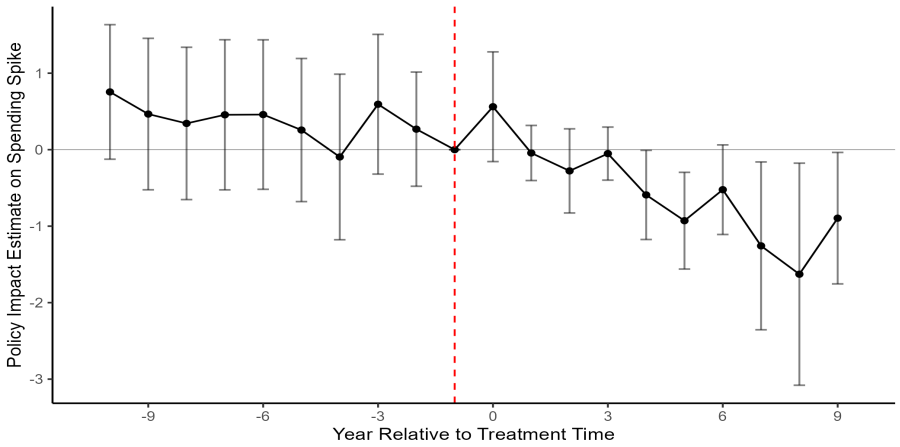


Figure 5: Staggered event study estimates for year-end spending surges after a rollover liberalization. Spending spikes are the ratio of last month spending relative to the average during the rest of the fiscal year. The unit of analysis is a purpose-specific expenditure item, there are year and unit fixed effects, and standard errors are clustered on the unit level. The dashed red line marks the pre-treatment year, which serves as the reference period. Staggered Difference-in-Difference estimates were obtained using the estimator in Sun and Abraham (2021).

there seems to be a reduction of about unity, which corresponds to a 25% decrease in the spike. I show in Appendix 5.6 that, consistently with the model's prediction, this reduction of the spike is driven by lower spending at the end of the fiscal year rather than heightened spending during its early periods.

This analysis also shows that it takes some time to transition to the new equilibrium, as the effect only shows up about four years after the intervention. This delayed response could also partially explain why Baumann (2019) does not find a significant impact of rollover liberalization on spending spikes.[30]

If we are to take the analysis in Figure 5 as an indicator of how restrictive the case-by case rollover decisions are, the optimal policy would be to grant somewhat more generous rollover rights to agencies, as the welfare maximizing policy in Figure 4 is associated with a reduction in the spending spike of about 50%.[31] This is coherent with the qualitative evidence from Proeller (2021), in which interviewed bureaucrats state that

[30]However, there are also other reasons for which a reduction in spending spikes due to the liberalization could remain undetected by the analysis in Baumann (2019). It relies on a before-after comparison using t-tests in which time-varying confounders may not be accounted for. Moreover, the sample size is extremely small with the largest group consisting of 13 and the smallest only of a single observation. Thus, the test is likely to be lacking in power and prone to type II error. Finally, the analyzed spending sequences are highly aggregated, such that agencies may not have discretion over the timing of many spending items contained therein (see Section 3.1).

[31]Which would be about twice as large as the impact measured in Figure 5.

rollover decisions are too restrictive to fully exploit their potential benefit. Thus, according to this logic, the current regime of limited rollover in Switzerland is located to the left of the optimum in Figure 4.

The results in Figure 5 also highlight a potential caveat of the welfare analysis in Section 3. Namely, in my model there is no such thing as a convergence period to the new equilibrium, such that spending behavior adjusts immediately when a new rollover rule is enforced. Thus, the welfare impact of the adaption periods in Figure 5 is not reflected in the calculations in Section 3. In consequence, the calculated results should be interpreted as gains once the new equilibrium has been reached. Having the model reflect a transition period as observed in Figure 5 would require a gradual learning mechanism in which the agents adapt to the rules of the new equilibrium (see e.g., Fudenberg and Levine 2009; Doraszelski et al. 2018). Even though such an adjustment channel is absent in my model, there is little reason to believe that it would overturn the welfare results from Section 3.3.

4 Conclusion

In this paper, I build a framework to analyze rollover policy in a multi-period budgeting principal-agent model. The key for efficiency gains through rollover policy is to achieve a more desirable sub-annual timing of spending for the principal, i.e., align the spending policy implemented by the agency with the principal's optimal solution. In this respect, especially the occurrence of spending spikes before the budget expires has been identified to cause inefficiencies (Liebman and Mahoney, 2017; Eichenauer, 2020; Klymak and Baumann, 2022). However, I show in Proposition 1 that heightened spending at the end of the year is a naturally occurring consequence of optimally insuring against uncertainty in a multi-period budget setting. Therefore, the inverse conclusion that a spending spike must necessarily be sub-optimal or that a policy evening out spending must be strictly beneficial cannot be made. In addition, I show that my model is able to capture the notion that restricting rollover may grant the principal more control over the agent, if their incentives are misaligned. Therefore, it can explain why parliaments generally are hesitant to give their agencies completely free rollover rights, even though this restricts their possibilities to save optimally in response to shocks. I calibrate this model to Swiss federal public consulting spending, which is an institutional framework that closely mirrors the setup of the model. I use this calibration to show that welfare gains are possible by allowing partial rollover. In this particular case, also taking opportunity costs and returned budget to the principal into account, it turns out that allowing agencies to retain about one third of unspent funds at the end of the fiscal year is the best rule, leading to welfare gains of just above 23% in terms of the Lucas number and 11% in terms of the compensating variation. Finally, I analyze a reform which took place in Switzerland liberalizing rollover at the federal level to confirm some of the model's core predictions

about the impact of rollover rights on the sub-annual timing of spending. The results imply that the reform led to welfare improvements, but is still somewhat too restrictive to reach the optimally achievable gains. Given their prevalence, we still have a surprisingly limited understanding of multi-period budgeting settings as well as their designs. Consequently, this paper leaves many relevant questions open to be addressed by future research. For example, the length of the budgeting horizon in a model of the type presented in this paper may interact with rollover policy in a way that choosing them jointly could lead to even larger welfare gains. Though such an analysis is outside the scope of this paper, it is an example of the many unexplored avenues left open to researchers interested in multi-period budgets.

5 Appendix

5.1 Lemma 1

Formally, Lemma 1 requires to show that $\{\lambda \in \mathbb{R}_+ | \mathcal{B} = 1\} \neq \emptyset$. To do this, note that the principal's optimality condition in period $\mathcal{M}$ states that:

$$\lambda = \beta \mathbb{E}\left[\frac{\partial}{\partial \mathcal{B}} v_P(\mathcal{B} + \mathbb{E}[k(\mathcal{R})], 1, z') | z\right]\left(1 + \frac{\partial \mathbb{E}[k(\mathcal{R})]}{\partial \mathcal{B}}\right) \tag{5}$$

Then note, that in this multi-period budget setting as long as the flow utility is convex and continuously differentiable, the standard proofs for concavity and differentiability of the value function in $\mathcal{B}$ from Stokey and Lucas Jr (1989) and Benveniste and Scheinkman (1979) still go through. This is the case if $u(\cdot)$ is concave, returns to scale in production are non-increasing and effort aversion is convex. Therefore $\mathbb{E}\left[\frac{\partial}{\partial \mathcal{B}} v_P(\mathcal{B} + \mathbb{E}[k(\mathcal{R})], 1, z') | z\right]$ is strictly decreasing in $\mathcal{B}$ such that $\mathbb{E}\left[\frac{\partial}{\partial \mathcal{B}} v_P(1 + \mathbb{E}[k(\mathcal{R})], 1, z') | z\right]$ exists and is unique.

Moreover, note that $\frac{\partial \mathbb{E}[k(\mathcal{R})]}{\partial \mathcal{B}}$ is going to be weakly positive, i.e., the amount of funds returned to the principal is non-decreasing in the allocated budget. To see this, consider the equilibrium in which the agency receives an annual budget of $\tilde{\mathcal{B}}$ and implements an optimal spending sequence $\{\tilde{x}\}_0^\infty$, which is associated with expected leftover funds at the end of the fiscal year $\tilde{\mathcal{R}}$. Now say the principal increases the budget to some $\hat{\mathcal{B}} > \tilde{\mathcal{B}}$. Assume that the associated spending sequence $\{\hat{x}\}_0^\infty$ is such that $\hat{\mathcal{R}} < \tilde{\mathcal{R}}$. This implies that the optimal spending policy under $\hat{\mathcal{B}}$ is lower than under $\tilde{\mathcal{B}}$ and also would have been feasible under $\tilde{\mathcal{B}}$, which contradicts the optimality of $\{\tilde{x}\}_0^\infty$. Therefore, $\frac{\partial \mathbb{E}[k(\mathcal{R})]}{\partial \mathcal{B}} < 0$ leads to a contradiction and we can conclude that the expected year-end reserves are weakly increasing in the annual budget.

Thus, λ is a product of two continuous $\mathbb{R} \to \mathbb{R}$ functions of $\mathcal{B}$, one of which is decreasing and one of which is increasing. This implies that also λ is a smooth $\mathbb{R} \to \mathbb{R}$ function of $\mathcal{B}$, such that there is a unique λ associated with normalizing $\mathcal{B} = 1$. Note however that, since λ may be non-monotonic in $\mathcal{B}$, the inverse is not true, i.e., there could be several optimal $\mathcal{B}$ consistent with a normalized λ. This is another reason for which I choose to use $\mathcal{B}$ rather than λ as the model's numéraire, deviating from Liebman and Mahoney (2017). Note however, that in the setup of Liebman and Mahoney (2017), this is a non-issue since the agency never returns any budget to the principal such that $\frac{\partial \mathbb{E}[k(\mathcal{R})]}{\partial \mathcal{B}} = 0$ always holds. Thus, in a setup without budgetary slack, it would be inconsequential whether one decides to normalize λ or $\mathcal{B}$.

5.2 Proposition 1

Consider a scenario in which the principal were able to implement her optimal spending decision herself given a budget $\mathcal{B}$ in a fiscal year with $\mathcal{M}$ sub-periods. For simplicity of notation, denote her flow utility $f(x) = u(y(x, z))$. In some period m, her first order condition for the level of saved budget is:

$$f'(x_m) + \lambda_m = \beta \mathbb{E}\left[\frac{\partial}{\partial B} v_p(B - x_m, m', z')|z\right] \tag{6}$$

In which λ_m is the Lagrange multiplier on the constraint capturing the shadow value of budget in period m and prime superscripts on variables denote their value next period.

Now assume that the principal would prefer (on average) even spending throughout the fiscal year, such that the optimal spending policy does not depend on the fiscal period. Therefore, $\mathbb{E}[f'(x_m)] = \mathbb{E}[f'(x_{\tilde{m}})]$ for any two periods m and $\tilde{m}$. Note that this also implies that $\mathbb{E}[v_p(B, m, z)] = \mathbb{E}[v_p(B, \tilde{m}, z)]$. To see this, remember that $v_p(B, m, z) = \mathbb{E}\left[\sum_{t=m}^{\infty} \beta^t f(x_t^*)|z\right]$, take unconditional expectations on both sides and apply the even spending assumption. This results in a term which is independent of m. By the envelope condition, this implies that also $\mathbb{E}[\lambda_m] = \mathbb{E}[\lambda_{\tilde{m}}]$.

Also note that Equation 6 can be written in semi-sequential form, writing out the value function up to the start of the next fiscal year explicitly:

$$v_p(B_t, m, z_t) = \mathbb{E}\left[\sum_{t=m}^{\mathcal{M}} \beta^{t-m} f(x_t^*) + \beta^{\mathcal{M}-m+1} v_p\left(B_t - \sum_{t=m}^{\mathcal{M}} x_t^* + \mathcal{B}, 1, z_{\mathcal{M}+1}\right) \middle| z_t\right]$$

Inserting this expression into Equation 6 and using the implications of the even spending assumption will result in a contradiction. After taking unconditional expectations, we get:

$$\mathbb{E}[f'(x_m^*)] = \mathbb{E}\left[-\lambda_m + \sum_{t=m'}^{\mathcal{M}} \beta^{t-m'} f'(x_t^*) + \beta^{\mathcal{M}-m+1} \frac{\partial}{\partial B} v_p(B_1, 1, z_1)\right] \tag{7}$$

In which $B_1 = B_t - \sum_{t=m}^{\mathcal{M}} x_t^* + \mathcal{B}$. However, if $\mathbb{E}[\lambda_m] = \mathbb{E}[\lambda_{\tilde{m}}]$, the first and the last term on the right-hand side amount to $(\beta^{\mathcal{M}-m+2} - 1)\mathbb{E}[\lambda_m]$, since the envelope condition states that $\lambda_m = \frac{\partial}{\partial B} v_p(B, 1, z)$. But this implies that we have $\mathbb{E}[f'(x_m^*)] = \mathbb{E}[\sum_{t=m'}^{\mathcal{M}} \beta^{t-m'} f'(x_t^*)] + (\beta^{\mathcal{M}-m+2} - 1)\mathbb{E}[\lambda_m]$, which clearly is incompatible with the even spending assumption that $\mathbb{E}[f'(x_m^*)] = \mathbb{E}[f'(x_{\tilde{m}}^*)]$ since the right hand side depends on m, even if expenditure and Lagrange multipliers are constant in expectation. Thus, the even spending assumption leads to a contradiction, such that $\mathbb{E}[\lambda_m] = \mathbb{E}[\lambda_{\tilde{m}}]$ and by extension the even spending assumption cannot hold.

Given the intuition described in Section 2, one would expect the principal's optimal spending policy to

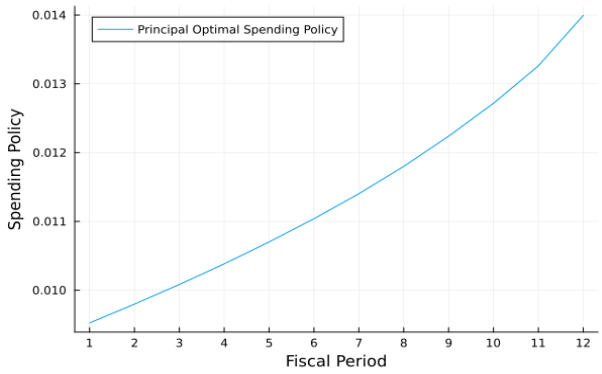


Figure 6: Optimal spending policy for the principal under full rollover given that $z_t = \mathbb{E}[z]$ and $B = 1$. The model parameters are set to the values calibrated in Section 3.

increase over the fiscal year, since the precautionary motive is stronger when the budget allocation period is still far away. In Figure 6 I show that this is indeed the case for the full rollover calibration obtained in Section 3. The displayed policy is for $z_t = \mathbb{E}[z]$ and $B = 1$. The rise in spending preference is quite pronounced, increasing by almost 50% from the first to the last fiscal period, even though rollover is completely free. Though this is a pronounced rise in spending preferences, it should be noted that for my calibration this rise would be even steeper for the agency under full rollover, such that the principal would still prefer more even spending than the agent throughout the year.

5.3 Proposition 2

I start by showing that departing from a full rollover equilibrium, constraining rollover will decrease savings and increase contemporaneous spending for the agency. Begin with the agency's first order condition for contemporaneous spending x_m for some $m < \mathcal{M}$:

$$f'(x_m) - e'(x_m) = \beta \mathbb{E}\left[\frac{\partial}{\partial B} v_A(B', m', z')|z\right] - \lambda_m^A \tag{8}$$

Note that the Lagrange multiplier for the constraint that $x_m \in [0, B]$ can take on a different value for the agency than for the principal, which is why λ_m^A carries the superscript here. Now assume that a new and more restrictive rollover rule is imposed, in the sense that $k(\mathcal{R}) \leq \mathcal{R}$ for all $\mathcal{R}$ and the inequality being strict for at least some $\mathcal{R}$. Denote the agency's value function under this constrained rollover regime $v_A^C(\cdot)$. Note that $v_A(B, m, z) - v_A^C(B, m, z)$ is non-decreasing in B. To see this, note that $v_A(\cdot)$ and $v_A^C(\cdot)$ are optimal value functions and any spending sequence which is feasible under the constrained rollover is also feasible under free rollover. Therefore, if moving from some budget B to $\tilde{B} > B$ were to increase $v_A^C(\cdot)$ more than $v_A(\cdot)$, this would contradict the optimality of $v_A(\cdot)$. Since both functions are continuously differentiable in

B and $v_A(B,m,z) - v_A^C(B,m,z)$ is weakly increasing in B, we also have that:

$$\frac{\partial}{\partial B} v_A(B,m,z) \geq \frac{\partial}{\partial B} v_A^C(B,m,z)$$

Moreover, note that from the derivation of the production function in Section 2 we have that $\lim_{x_m \to 0} f'(x_m) = \infty$. Moreover, it is natural to assume that $\lim_{x_m \to 0} e'(x_m) < \infty$. Therefore, we also have that $\lim_{B \to 0} \frac{\partial}{\partial B} v_A(B,m,z) = \infty$. Given that this inada condition holds for both the flow utility and the value function, we can rule out corner solutions for any period $m < \mathcal{M}$ such that $\lambda_m = 0$ during the fiscal year.[32] Therefore, moving to the constrained rollover regime is going to decrease the right-hand side of Equation 8. Since $f(\cdot)$ is concave and $e(\cdot)$ is convex, the left-hand side of Equation 8 is concave such that x_m has to increase in response.

What is left to show is that the larger $e'(\cdot)$, the more situations there will be in which $x_A^* < x_p^*$ such that marginally increasing x_A^* will align the two. To show this, it suffices to prove that x_A^* is decreasing if marginal effort aversion shifts upward, since the principal's preferences are characterized by zero effort aversion.[33] To this end, assume that there were some parameter δ in which marginal effort aversion is increasing for any $x_m > 0$. For the sake of concise notation again denote the flow utility by $f(\cdot)$ and the agency's discount factor for period t by D_t. Then, one can write the agency's problem as a sequence of choosing a sequence of available budgets in every period:

$$v_A(B,m,z) = \max_{\{B_t\}_{t=0}^{t=\infty}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} D_t (f(B_t - B_{t+1}) - e(B_t - B_{t+1})) \right]$$

Subject to the budget renewal process every $\mathcal{M}$ periods. Taking the associated Lagrangian's first order condition with respect to B_{t+1} gives: $f'(B_t^* - B_{t+1}^*) - e'(B_t^* - B_{t+1}^*) = \lambda_t + \beta \mathbb{E}_0[f'(B_{t+1}^* - B_{t+2}^*) - e'(B_{t+1}^* - B_{t+2}^*)]$, in which λ_t is the Lagrange multiplier on the budget constraint, i.e. the marginal value of budget in period t. When taking the total derivative of this condition with respect to δ, by definition of δ we know that the derivative of the first and the last term must be negative. Therefore, we have that $\frac{\partial}{\partial \delta} \lambda_t \geq \frac{\partial}{\partial \delta} e'(x_t^*)$, though both terms are smaller than zero. Now we can go back to the first order condition in Equation 8 and apply the implicit function theorem to get:

$$\frac{\partial x_m}{\partial \delta} = - \frac{\frac{\partial}{\partial \delta} \mathbb{E}\left[f'(x_m) - e'(x_m) - \beta \frac{\partial}{\partial B} v_A(B', m', z') \right]}{f''(x_m) - e''(x_m) + \beta \mathbb{E}\left[\frac{\partial^2}{\partial B^2} v_A(B', m', z') \right]} \tag{9}$$

Using the envelope condition and that $\frac{\partial}{\partial \delta} \lambda_t \geq \frac{\partial}{\partial \delta} e'(x_t^*)$ in the numerator of the fraction in Equation 9,

[32] In the last period of the fiscal year, there could be a positive Lagrange multiplier though, which is why the proposition only applies for the first 11 months of the fiscal year.

[33] Note that this does not imply that the agency always spends less than the principal would prefer since they also differ in their discounting pattern. It does however imply, that the subset of the state space for which $x_A^* < x_p^*$ increases with marginal effort aversion.

we get that it must be < 0. Since the value function is concave, effort aversion is convex and also the flow utility from production is concave, the denominator must also be negative. Therefore, the overall expression is negative and an upward shift in marginal effort aversion must decrease contemporaneous spending.

5.4 Ratchet Effect & Limited Commitment

Rathet Effect. If unspent budget sent a signal to the principal in response to which she would decrease the budget, the baseline calibration to $k(\mathcal{R}) = 0 \, \forall \, \mathcal{R}$ would potentially be misleading.

	Dependent variable:		
	Allocated Consulting Budget		
	(1)	(2)	(3)
Unspent Budget $(t-1)$	−187,700	204,916	225,490
	(465,158)	(199,535)	(204,816)
Agency fixed effects	No	Yes	Yes
Year fixed effects	No	No	Yes
Observations	236	236	236
R^2	0.001	0.902	0.904

*p<0.1; **p<0.05; ***p<0.01

Table 2: This table documents the absence of a ratchet effect in the calibration data. An observation corresponds to one agency-fiscal-year pair. Standard errors are reported in parentheses.

If this were the case, we would expect to observe a negative relationship between year-end leftover funds and the following year's budget. Table 5.4 documents that my calibration data does not exhibit such a relationship. The unspent budget variable is calculated as one minus the share of budget spent in a given fiscal year. This variable is lagged by one year since the leftover funds in year x would determine the budget in year $x + 1$ if a Ratchet effect were present. Model one does not include any fixed effects, model two only agency fixed effects and model three includes both agency and year fixed effects. None of the three show a significant relationship at any conventional level of statistical significance, lending credibility to the calibration to the zero rollover parameterization.

Limited Commitment. I conduct a similar test as above for the possibility of limited commitment. As outlined in the main text, it would imply a negative correlation between realizations of $k(\mathcal{R})$ and the following year's budget for an agency. Note that this is different from the test above, in which the independent variable was leftover budget rather than rolled over budget. The reason for this is that, as opposed to a Ratchet effect, with limited commitment, the principal would not necessarily react to high leftover funds if they are not rolled over.

	Dependent variable:		
	Growth rate of $\mathcal{B}$ following year		
	(1)	(2)	(3)
Realization of $k(\mathcal{R})$	−3.513e-10	−8.389e-11	4.675e-11
	(5.173e-10)	(9.206e-10)	(9.447e-10)
Agency fixed effects	No	Yes	Yes
Purpose fixed effects	No	No	Yes
Observations	236	236	236
R^2	0.002	0.190	0.198
Note:			*p<0.1; **p<0.05; ***p<0.01

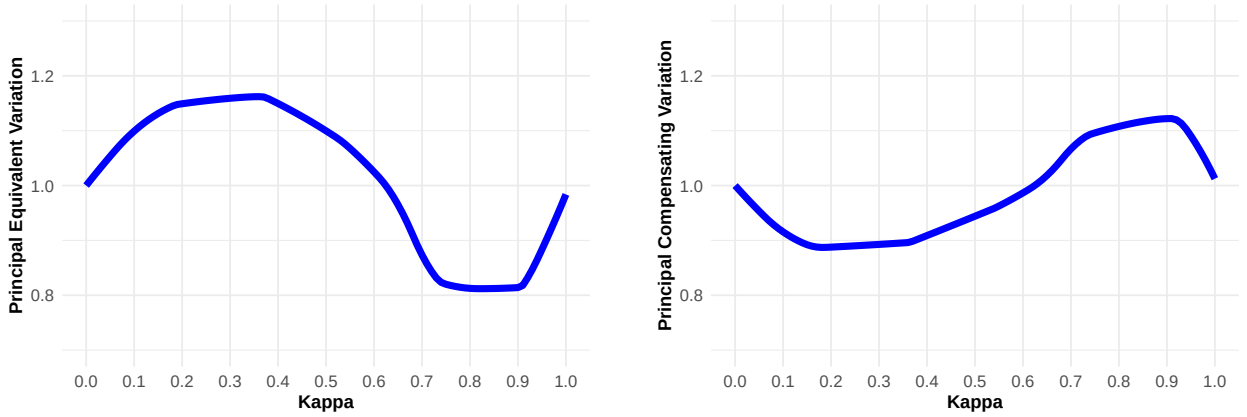
Table 3: This table shows that the data does not feature a measurable correlation between the reserves built by an agency and the growth rate of its budget in the following year. This is evidence for the principal being able to commit to a budget rule credibly.

The analysis in Table 5.4 shows that such a correlation is absent in my budgeting data across all fixed-effect specifications. Therefore, I conclude that the principal is able to credibly commit to a budget and refrain from raiding the agency's rolled over funds. In consequence, I impose $\pi = 1$ in the calibration in Section 3, such that the displayed counterfactual equilibria are actually implementable by the agency.

5.5 Alternative Welfare Measures

In Section 3.3, I report welfare gains in terms of the Lucas consumption equivalent, which is commonly used in dynamic macroeconomic models. For completeness and comparability to other work however, I also calculate the compensating and equivalent variation in a similar fashion as Liebman and Mahoney (2017) quantifies the compensating variation. The equivalent variation gives the factor by which the agency's budget has to be scaled, while keeping the principal's expenses fixed, in the baseline ($\kappa = 0$) equilibrium to make the principal indifferent to the equilibrium under an alternative rollover rule. Similarly, the compensating variation scales the agency's budget under some equilibrium to make the principal indifferent to the baseline use-it-or-lose-it rollover rule. Those applications are extrapolations of these welfare concepts as they are typically used in comparative static settings, but still informative to put the welfare gains reported in Section 3.3 into context.

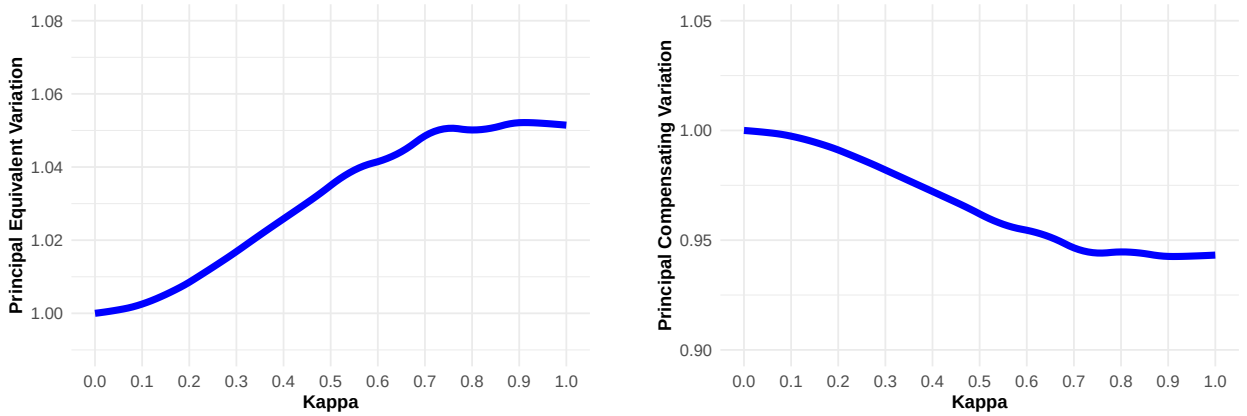
Similar plots as in Figure 4a from the main text are displayed in Figure 7 for the $k(\mathcal{R}) = \kappa\mathcal{R}$ rule. The line-plot on the left shows the compensating and the right the equivalent variation. Note that, for the compensating variation, numbers below one correspond to welfare improvements relative to the use-it-or-lose-it rule since it implies that budget has to be scaled down for the principal to be indifferent to it. The pattern of the equivalent variation in Figure 8 is qualitatively very similar to the Lucas number reported in the main text. Quantitatively however, the gains are about 30% smaller whereas the losses of liberalizing



(a) Equivalent Variation

(b) Compensating Variation

Figure 7: Welfare metrics for the $k(\mathcal{R}) = \kappa\mathcal{R}$ policy which allows to roll over a fraction of year-end savings.



(a) Equivalent Variation

(b) Compensating Variation

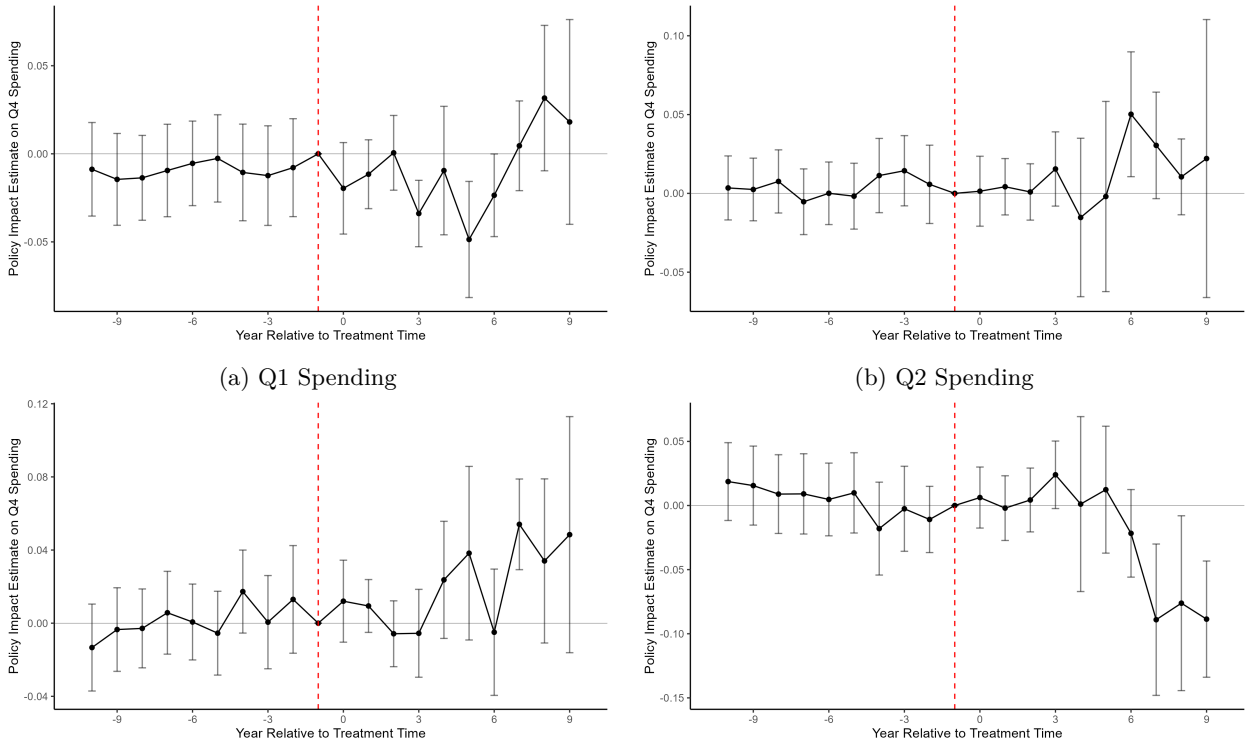
Figure 8: Welfare metrics for the $k(\mathcal{R}) = \min\{\mathcal{R}, \kappa\mathcal{B}\}$ policy which allows to roll over up to a fraction of the allocated budget.

rollover too much are of a similar magnitude. The welfare gains at the optimum (allowing to roll over about one third of unspent budget) remain quite substantial, at around 16%. In terms of the compensating variation, these gains turn out to be even somewhat smaller at around 11%. Moreover, as measured by the equivalent variation, the same gains can be achieved with any $\kappa \in [0.15, 0.35]$.

For the $k(\mathcal{R}) = \min\{\mathcal{R}, \kappa\mathcal{B}\}$ rollover rule, the patterns for the compensating and equivalent variation turn out to be very similar to the Lucas consumption equivalent reported in Figure 3. Similarly to the analysis in the main text, the gains plateau around $\kappa = 0.7$ and steadily rise before that. However, also for this policy the achievable gains are quantitatively somewhat smaller than for the Lucas number (by about 3%).

5.6 Robustness of Reduced Form Evidence

The simulations in Section 3.3, which are associated with a similar reduction in the spending spike as the estimates displayed in Figure 5 also make different predictions with respect to the distribution of sub-annual spending. Specifically, the reduction in the spending spike should mostly be driven by a reduction of year-end spending rather than an increase in earlier periods. Running event studies similar to the one displayed in



(a) Q1 Spending

(b) Q2 Spending

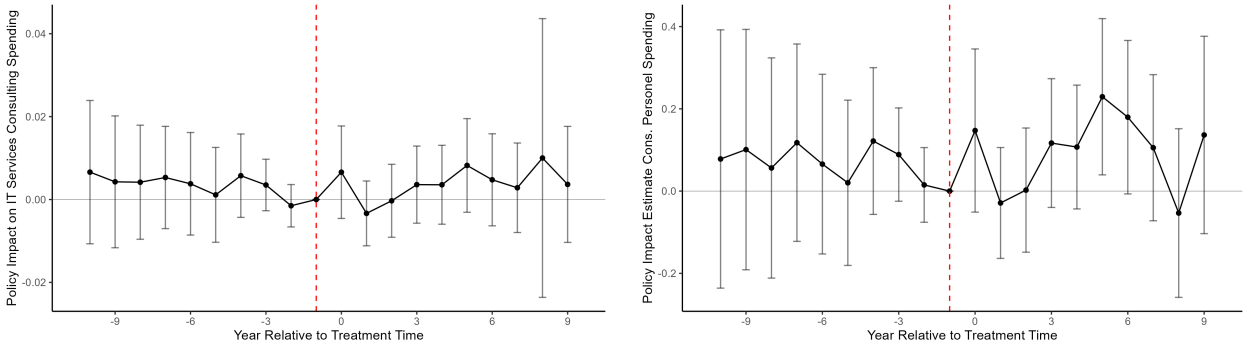
(c) Q3 Spending

(d) Q4 Spending

Figure 9: Event studies for quarterly spending in response to rollover reform. The unit of analysis is an expenditure item over time. All estimates are obtained using the staggered DiD technique from Sun and Abraham (2021) and include time and unit fixed effects. Standard errors are clustered on the unit level.

Figure 5, I obtain the reform's impact on the share of budget spent in each quarter of the fiscal year. The resulting estimates are displayed in Figure 9. In accordance with the simulated model's prediction, there is no measurable response of spending during the first three quarters. Each panel has one or two significant point estimates, but this is to be expected considering that there are 18 point estimates per panel. Only during the first quarter, there are two point estimates which are significant, but they appear more random rather than indicative of systematically reduced spending. During the last quarter on the other hand, a pronounced and persistent reduction in spending can be observed in the periods for which the spending spike in Figure 5 decreases. This supports the model's prediction that the smoothing out effect of rollover comes mostly from reducing year-end spending as its impact on early spending is negligible.

As mentioned in the main text, one threat for the reduced-form analysis is that the relative allocation of



(a) IT service consulting costs

(b) Consulting personnel costs

Figure 10: Event studies for the share of funds allocated to IT service consulting and consulting personnel costs. The unit of analysis is an expenditure item over time. All estimates are obtained using the staggered DiD technique from Sun and Abraham (2021) and include time and unit fixed effects. Standard errors are clustered on the unit level.

funds across spending items may have changed with the reform, which could also impact the sub-annual distribution of spending. Only two of the spending items included in the calibration and estimation data feature enough observations across all periods to test this prediction, but for those I show in Figure 10 that the share of the overall budget allocated to them has not changed in response to the reform. Again, there is one significant point estimate for the right-hand side panel, but this is still very likely under the null-hypothesis of no impact given that there are almost 20 point estimates in the event study. This robustness check lends additional credibility to the results in Section 3.4.

References

S. R. Aiyagari. Uninsured idiosyncratic risk and aggregate saving. The Quarterly Journal of Economics, 109 (3):659–684, 1994.

G. A. Akerlof. Procrastination and obedience. American Economic Review, 81(2):1–19, 1991.

P. Bajari and S. Tadelis. Incentives versus transaction costs: A theory of procurement contracts. RAND Journal of Economics, pages 387–407, 2001.

O. Bandiera, A. Prat, and T. Valletti. Active and passive waste in government spending: evidence from a policy experiment. American Economic Review, 99(4):1278–1308, 2009.

S. Baumann. Putting it off for later: Procrastination and end of fiscal year spending spikes. The Scandinavian Journal of Economics, 121(2):706–735, 2019.

L. M. Benveniste and J. A. Scheinkman. On the differentiability of the value function in dynamic models of economics. Econometrica, pages 727–732, 1979.

M. C. Best, J. Hjort, and D. Szakonyi. Individuals and organizations as sources of state effectiveness. American Economic Review, 113(8):2121–2167, 2023.

E. Bosio, S. Djankov, E. Glaeser, and A. Shleifer. Public procurement in law and practice. American Economic Review, 112(4):1091–1117, 2022.

R. Chetty. A general formula for the optimal level of social insurance. Journal of Public Economics, 90 (10-11):1879–1901, 2006.

R. Crawford, C. Emmerson, and G. Tetlow. A survey of public spending in the uk. IFS Briefing Note BN43. London, United Kingdom: Institute for Fiscal Studies, 2009.

U. Doraszelski, G. Lewis, and A. Pakes. Just starting out: Learning and equilibrium in a new market. American Economic Review, 108(3):565–615, 2018.

A. S. Dunk and H. Nouri. Antecedents of budgetary slack: A literature review and synthesis. Journal of Accounting Literature, 17:72, 1998.

V. Eichenauer. December fever in public finance. Technical report, KOF Working Papers, 2020.

M. Elmassri and E. Harris. Rethinking budgetary slack as budget risk management. Journal of Applied Accounting Research, 12(3):278–293, 2011.

X. Freixas, R. Guesnerie, and J. Tirole. Planning under incomplete information and the ratchet effect. The Review of Economic Studies, 52(2):173–191, 1985.

D. Fudenberg and D. K. Levine. Learning and equilibrium. Annu. Rev. Econ., 1(1):385–420, 2009.

U. General Accounting Office. Federal year-end spending: Symptom of a larger problem. 1980.

U. General Accounting Office. Year-end spending reforms underway but better reporting and oversight needed. 1998.

F. Khalil, D. Kim, and J. Lawarree. Use it or lose it. Journal of Public Economic Theory, 21(6):991–1016, 2019.

M. Klymak and S. Baumann. Paying over the odds at the end of the fiscal year. evidence from ukraine. 2022.

J. B. Liebman and N. Mahoney. Do expiring budgets lead to wasteful year-end spending? evidence from federal procurement. American Economic Review, 107(11):3510–3549, 2017.

R. E. Lucas. Models of business cycles, 1987.

R. E. Lucas. Macroeconomic priorities. American Economic Review, 93(1):1–14, 2003.

C. P. McCue, E. Prier, and R. J. Lofaro. Examining year-end spending spikes in the european economic area: a comparative study of procurement contracts. Journal of Public Budgeting, Accounting & Financial Management, 33(5):513–532, 2021.

M. F. McPherson. An analysis of year-end spending and the feasibility of a carryover incentive for federal agencies. 2007.

I. Proeller. Neues führungsmodell für die bundesverwaltung: Vollzugs-und wirkungsevaluation. 2021.

P. Rey and B. Salanie. On the value of commitment with asymmetric information. Econometrica, pages 1395–1414, 1996.

W. P. Rogerson. Simple menus of contracts in cost-based procurement and regulation. American Economic Review, 93(3):919–926, 2003.

D. E. M. Sappington. Incentives in principal-agent relationships. Journal of Economic Perspectives, 5(2): 45–66, 1991.

C. Siemroth. Ending wasteful year-end spending: On optimal budget rules in organizations. International Economic Review, 2022.

N. L. Stokey and R. E. Lucas Jr. Recursive methods in economic dynamics. Harvard university press, 1989.

L. Sun and S. Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics, 225(2):175–199, 2021.

D. Tarschys. Time horizons in budgeting. OECD Journal on Budgeting, 2(2):77–103, 2003.

J. Tirole. Procurement and renegotiation. Journal of Political Economy, 94(2):235–259, 1986.

R. A. Webb. The impact of reputation and variance investigations on the creation of budget slack. Accounting, Organizations and Society, 27(4-5):361–378, 2002.