INTERNATIONAL ECONOMIC REVIEW
Vol. 0, No. 0, December 2023 DOI: 10.1111/iere.12684

ENDING WASTEFUL YEAR-END SPENDING: ON OPTIMAL BUDGET RULES IN ORGANIZATIONS*

By Christoph Siemroth

University of Essex, U.K.

What can organizations do to minimize wasteful year-end spending? I introduce a two-period model to derive optimal budget roll-over and audit rules. A principal tasks an agent with using a budget to fulfill the organization's spending needs, which are private information of the agent. The agent can misuse funds for private benefit. The optimal rules allow the agent to roll-over a share of the unused funds, but not necessarily the full share, and in most cases to audit only sufficiently large spending. The optimal audit rule can change once fund roll-over is allowed. Strategically underfunding the agent can be optimal.

1. INTRODUCTION

Many organizations or divisions around the world are granted annual budgets by their principals—which expire at the end of the fiscal year—to fulfill their mission.[1] These expiring budgets induce agents to "use or lose" these funds in the last month or week of the fiscal year,[2] which can result in spending with little value to the organization. In a survey among U.S. DoD staff responsible for spending, 95% said there was an efficiency problem with year-end spending (McPherson, 2007). On average, interviewees estimated that 32% of year-end spending was on low priority items or at least partially wasted.

This article investigates what other rules could be adopted to incentivize agents to use funds more efficiently. Liebman and Mahoney (2017) show that allowing full roll-over of unused funds to the next year can alleviate wasteful year-end spending. But they have not studied how much roll-over is optimal, nor have they studied the interaction of roll-over rules and auditing. In a new model with these aspects, I contribute three key findings. First, allowing partial roll-over, where only a fraction of funds is rolled over to the next year, can be better than full roll-over. This is because partial roll-over can in some circumstances induce more saving—and hence less wasteful spending—by the agent. Second, absent fund roll-over, the

*Manuscript received July 2022; revised November 2023.

I thank Yair Antler and Albin Erlanson for thoughtful discussions on this topic, as well as Stuart Baumann, Pierre Boyer, Pierre-Edouard Collignon, Evgeniya Dubinina (discussant), Daniel Garrett, Jayant Ganguli, Michael Gibbs, Piero Gottardi, Hans Peter Grüner, Yaron Leitner, Olivier Loisel, Felix Schleef, Sergey Stepanov, Wim van der Stede, and audiences at Essex, Mannheim, St. Gallen, École Polytechnique, and the ZEW Public Finance and SAET 2023 Paris conferences for useful comments. I also thank the editor Masaki Aoyagi for his patience and guidance, and three referees for their help in improving the manuscript. This work was supported by the Economic and Social Research Council [grant number ES/T006048/1]. Please address correspondence to: Christoph Siemroth, Department of Economics, University of Essex, Wivenhoe Park, CO4 3SQ Colchester, U.K. Phone: +44 (0) 1206 873518. E-mail: *christoph.siemroth@essex.ac.uk*.

[1] A public-sector example is U.S. Congress (principal), which grants funds and tasks the Department of Defense (DoD) (agent) with defending the nation. In the private sector, CEOs give their heads of marketing a budget to promote the firm brand. In universities, deans or grant funders give funds to academics for research, staff, and conferences.

[2] Year-end spending sprees are well documented in the public sector (e.g., Liebman and Mahoney, 2017; Baumann, 2019 and references therein), and anecdotes suggest the phenomenon is also common in parts of the private sector (e.g., Digiday, 2017 for marketing).

optimal audit rule should only audit large spending amounts. But, third, if fund roll-over is allowed, then the optimal audit rule can change its structure, so that only spending amounts in the interior of the budget set should be audited, not those at the upper end. Consequently, auditing is not something that can be investigated independently of roll-over rules, as these very much affect the way auditors should be deployed.

The status quo with expiring budgets has several undesirable consequences. First, the agent might hastily spend the remaining funds on low value items (e.g., Liebman and Mahoney, 2017), because there is no useful purpose left at year-end, or there is no time to execute the spending well (e.g., because the best contractor is unavailable on short notice). Second, the agent might spend the funds on currently unneeded but durable "assets," in the hopes the expiring budget can still yield a benefit in the future. For example, Hurley et al. (2014) report a case where a military officer was ordered to buy a train-wagon-load of toilet paper at year-end. While such tricks might not be a complete waste of funds, toilet paper as a currency is less fungible and storable than money, and hence loses value. Third, the agent might misuse the funds for personal gain and little value to the principal. An example might be a $9,000 chair bought by the Pentagon at year-end (Military Times, 2019), where one would think a $5,000 chair would have done as well. Another example is gadgets bought to satisfy the curiosity of a tech nerd working in an organization's IT department without benefit to the organization. The model in this article is motivated by this third interpretation, but is also consistent with the first and second.

In practice and in the model of this article, the source of the principal–agent conflict is that the principal does not know the agent's exact spending needs, so grants the agent some discretion in spending. The agent does not know of better uses for unused funds outside of his agency, nor does he take into account the cost of these funds like the principal does. Hence, the agent spends everything even if not all is needed to fulfill the principal's task.[3]

This article models the principal–agent interaction as a game of asymmetric information. The model has two years, and the agent receives an exogenous budget in each of them, which expires at year-end. The agency's spending need θ_t in year t is a random draw from a continuous distribution. This captures that future spending needs are uncertain, for example, it is unclear how many computers will break down in the agency's office during the coming year. The realization θ_t is privately observed by the agent in year t, who then decides on a spending amount for that year, subject to the budget constraint. Any spending up to the spending need θ_t fulfills those spending needs and generates a high value for the principal, but any spending above θ_t is a misuse of funds, as it yields no value to the principal. The agent receives a high marginal value from fulfilling spending needs, and a lower but positive marginal value from spending more. Hence, the agent wants to fulfill his mission first, but also values the $9,000 office chair, whereas the principal does not.

Without further additions, this model generates a year-end spending surge as observed in practice: The agent rationally spends everything and misuses funds, even if the spending need is considerably below the budget. This is a poor outcome for the principal, as any spending above need generates no value but generates costs (e.g., credit costs).

To investigate measures to mitigate this waste of funds, the principal commits to the following rules before the agent moves. First, the principal sets a roll-over rule $\Delta \in [0, 1]$, the share of unused funds that will be rolled over to the second year, whereas share $1 - \Delta$ is returned to the principal. Hence, for any unused dollar at the end of the first year, Δ dollars are added

[3] At least two more motives for excessive year-end spending are mentioned in the literature. First, the agent's fear to have his budget cut next year if not all funds are used, which is known as the ratchet effect (e.g., Freixas et al., 1985). Second, U.S. politicians appear to see unused funds as the agent not doing his job, applying pressure to spend, and thus providing a disincentive for agents to use their funds efficiently. Although this may sound silly to economists—viewing more spending as desirable irrespective of the value it generates or its opportunity costs—it may be rational from a political economy point of view. As unused federal funding reverts to the treasury only after five years, the only way to get a benefit for the constituency in the current four-year term is to pressure agencies to spend their annual budgets (e.g., McPherson, 2007).

to the agent's budget in the second. Second, the principal sets an audit rule for each year, and these rules specify for which spending amounts a *costly* audit will be triggered. An audit is the only way for the principal to observe the spending need θ_t and thus to identify fund misuse. The agent is punished if he misused funds and was audited.

Based on this model, I find that the optimal roll-over rule features some, but not necessarily full roll-over. That is, a positive share of the unused funds should be available to the agent next year, but surprisingly, a partial roll-over can be superior to a full roll-over under some conditions. The possibility of fund roll-over gives the agent a reason to save instead of misusing leftover funds. But so far there has been no analysis of an optimal roll-over rule, nor has partial roll-over been suggested as preferable to full roll-over. Partial roll-over can be optimal if the agent receives little value from fund misuse and if the principal's audit and fund costs are high. This finding contrasts with the current rules in many public sectors such as in the United States, Canada, or Germany, where budget roll-over is typically not possible.

The intuition why partial roll-over can be optimal is as follows: Suppose there is no auditing. If there are very few spending needs in year 1, then if all unspent funds are rolled over ($\Delta = 1$), the agent would only save some for roll-over and misuse the rest. This is because the agent needs only so many additional funds to cover most expected spending needs next year. But if the roll-over rule "taxes" the roll-over ($\Delta < 1$), then it forces the agent to save more in order to have the same budget next year. This additional saving crowds out fund misuse. Consequently, "taxing" the roll-over can be optimal. In practice, it would also address concerns that agents accumulate too many savings over time.

Given large enough punishment, audits happen in equilibrium, but any agent who is audited has legitimately spent a lot due to a large spending needs realization, so no agent is punished in equilibrium. If the roll-over rule is not effective, that is, never induces the agent to save unused funds for roll-over, then the optimal audit rules in both years are threshold rules. A threshold rule audits all yearly spending above a threshold (which may differ by year), but does not audit below the threshold. There is still some scope for fund misuse for agents in years with low spending needs while staying under the audit threshold.

Under the optimal threshold audit rules, the principal tends to audit more (i.e., smaller spending amounts) the lower the cost of auditing and the larger the cost of funds. A larger cost of funds implies a larger loss for the principal from fund misuse, hence she is willing to audit more to prevent misuse. Moreover, if the annual budget is enough to cover all potential spending needs, then the optimal audit thresholds for both years are identical, otherwise the principal tends to audit more in the first than in the second year. This is because auditing in the first year not only discourages fund misuse, but also leads to more fund roll-over, which helps to satisfy more spending needs in the second year. For small budgets, the principal tends not to audit at all, because there is little scope for fund misuse, as the probability of spending needs below budget is small. For larger budgets, the principal tends to audit as long as audit costs are not too large or the cost of funds is large enough.

There is one exception where the audit rule is not a threshold rule in the first year, if the roll-over rule is effective in inducing the agent to save funds for roll-over. In this case, the agent does not spend his entire budget in the first year—unless the spending needs of the agency require it—even if there is no auditing, because he wants to have more funds available next year. The optimal audit rule in this case is an interval rule, which audits spending amounts just above and below what a saving agent would spend absent auditing. Consequently, the optimal interval rule audits spending amounts in the interior of the budget set, and might not audit the very largest spending amounts close to the total budget, where the roll-over incentive prevents fund misuse. Consequently, auditing is not something that can be investigated independently of roll-over rules, as these very much affect the way auditors should be deployed.

In an extension, I endogenize the amount of the annual budget. If the ratio of audit costs to cost of funds, and the cost of funds, are large enough, then it is optimal for the principal to grant a budget that is smaller than the maximum possible spending need realization. This implies there is a positive probability the agent cannot meet all spending needs. Still, this is better for the principal than a large budget with a high chance of fund misuse, as auditing is too costly. If endogenous budgets can be different every year, then the extensions indicate that budgets should weakly decrease over time: The agent has time to save to offset the budget decreases, and the decreasing budget strengthens incentives not to misuse funds.

The policy recommendations to reduce wasteful year-end spending are as follows: First, rolling over unused funds should be allowed. The worst that can happen is that funds are misused next year instead of this year, which is no worse than the status quo with use-it-or-lose-it behavior. Moreover, while auditing requires additional manpower, allowing fund roll-over is relatively inexpensive. Second, auditing and punishing fund misuse can help for sufficiently small audit costs. Usually only large spending amounts should be audited, but not small spending amounts far below the budget, which indicate funds were likely used in the principal's interest. If the roll-over rule is effective in inducing savings, then auditing the very largest spending amounts could be unnecessary. If audit costs are large or the cost of funds are very small, then no auditing is optimal. Third, setting a low budget—so not all spending needs can be fulfilled with positive probability—can be optimal if the audit costs and the cost of funds are large, or the agent's mission is of low priority.

1.1. *Literature.* This article contributes to the year-end spending literature. Liebman and Mahoney (2017) empirically document U.S. federal procurement year-end spending surges, and that IT spending made in the last week of the fiscal year is of lower quality than usual. They also show that allowing some roll-over of unused funds to the next year reduces year-end spending surges. Liebman and Mahoney (2017) also develop a model in which a precautionary savings motive for the agent generates more spending of less value at year's end. In their model, allowing roll-over can increase welfare. In this article, I contribute the first analysis of the optimal roll-over rule in this context as well as an analysis of optimal audit rules, whereas there is no auditing in their model. Major new insights are that the optimal roll-over rule may not feature full roll-over as they consider, and that the structure of the optimal audit rule changes in the presence of fund roll-over.

Baumann (2019) also empirically documents year-end spending surges in UK governmental agency spending, and investigates whether a precautionary savings motive is causing them. He finds that an alternative explanation—that agencies procrastinate in spending their funds—also plays a role. Hurley et al. (2014) and Brimberg and Hurley (2015) show that a rational risk-neutral planner aiming to maximize the value of spending will—due to the uncertainty of future spending needs—be conservative in spending early and often have unused funds at the end of the year. They also show that pressure to minimize unused funds at the end of the year leads to lower value spending. Their studies are critiques of expiring budgets, but their focus is not on analyzing alternatives, which is what I add in this article.

My model, and those in the year-end spending literature, take it as given that the agent receives a budget and some discretion in spending it, and investigate improvements within that structure. This should make the recommendations derived from this model easier to implement in practice. Malenko (2019) takes a step back and asks whether agents should receive budgets in the first place, or whether spending should get micromanaged more and funded by the principal directly. In his environment, there is a principal–agent conflict because the agent has preferences for overspending. The optimal mechanism separates small and large investments, so that small investments are funded at the agent's discretion from his budget, which is replenished over time. Large investments are funded by the principal directly, but only after an audit confirms it is worth to be implemented. If the large project is not worth it, the agent is punished for recommending it. Interestingly, this optimal mechanism does not have annual budgets; in fact, funds never expire. Hence, it could be viewed as an annual budget with complete and indefinite roll-over.

A related literature in corporate finance asks how headquarters should allocate internal funds to lower-level managers, when these managers have preferences for overspending but superior information about investments. Harris and Raviv (1996) show how optimal incentive budget allocations consists in capital spending limits and a procedure to relax them if warranted. Harris and Raviv (1998) extend that model if the manager chooses between multiple projects, and show that similar capital spending limits are optimal. Whereas the first paper is not dynamic, the second one has a dynamic interpretation, showing that full roll-over is optimal. But due to the difference in settings, in particular because the agent knows future realizations for certain, there is no partial roll-over. Roper and Ruckes (2012) analyze a setting where the manager has better information about future investment opportunities. They show that commitment to a rule, where funding now is tied to lower probability of funding later, can make the manager reveal their superior information and thus improve intertemporal capital allocation. Overall, these seminal papers are positive theories that rationalize observed budget procedures as optimal responses to agency conflicts, whereas the current article is more normative in that it models the observed status quo and asks which rules improve outcomes. This literature is valuable in explaining the existence of spending delegation, budget limits, and explaining the budget allocations. But there is no discussion of year-end spending, no interaction between audits and roll-over rules, and except for Harris and Raviv (1998), there is no discussion of budget roll-over, which this article adds.

Khalil et al. (2019) depart from the previous year-end spending literature by adding an agent effort choice, which endogenizes the cost of spending. They show that rigid budget policies without roll-over can have benefits—inducing more effort, lower costs, and more production—compared to roll-over policies, so that the principal sometimes prefers no roll-over. Their model thus provides a rationale for denying roll-over. In their model, the entire budget is always spent, so it is not designed for the analysis of curbing wasteful year-end spending.

Bird and Frug (2019) investigate how an agent can be incentivized to provide effort to implement investments, which are beneficial to the principal but costly to the agent. Rewards are costly to the principal but beneficial to the agent. While their model is not explicitly about budgets or year-end spending, it might give wasteful year-end spending a new interpretation as rewards for previous agent performance.

The auditing part of my model is related to the theoretical costly state verification literature, such as Townsend (1979), Border and Sobel (1987), Ben-Porath et al. (2014), and Li (2020). A new insight in the budget context is that allowing fund roll-over can make auditing unnecessary, or reduce the need for auditing. Moreover, the optimal audit rule changes depending on the effectiveness of the roll-over rule.

2. THE MODEL

2.1. *Setup.* There are two risk-neutral expected utility maximizing players, the principal (she) and the agent (he). This setting applies to organizations where an agent is given a budget to fulfill a mission, with some leeway in spending it.[4]

Time. Time is discrete. There are two periods in which the agent moves, which I call years, $t = 1, 2$. A minimum of two years is needed to investigate whether allowing to roll-over funds to the next year can mitigate inefficient year-end spending.[5] The year is the time frame for which a budget is granted to be spent by the agent, and in practice many organizations grant budgets for one year. The principal moves in $t = 0$.

[4] This is relevant in many public-sector organizations. In private-sector organizations, it applies to divisions which accountants call "cost centers," which means they incur costs to do a job but do not generate revenues themselves. This happens, for example, in IT, marketing, or R&D divisions.

[5] An earlier version of this model additionally divided the year into two subperiods (the last to be interpreted as year-end), each with an independent spending need realization. Adding these subperiods yielded the same spending and fund misuse decisions as this model, because it was a weakly dominant strategy for the agent to postpone fund misuse to year's end (exactly as observed empirically). To ease the exposition, I dropped this extra division into subperiods. Still, for this reason, fund misuse in this model can be interpreted as year-end spending.

State of the world: Spending needs. In every year, a random variable θ_t realizes, which is identically and independently distributed according to a continuous uniform distribution on $\Theta_t = [0, u]$, with $u > 0$. Let the density function be g and the cumulative distribution function (CDF) be G.[6] The agent observes the realization θ_t in year t, hence does not know realization θ_2 in year 1 (i.e., the agent cannot foresee the future). Moreover, θ_t is private information of the agent, so the principal does not observe it. θ_t is the *spending need* arising in year t in the agent's division or agency. For example, the more computer hardware breaks down and needs to be replaced, or the more temporary hires need to be made due to absences, the larger the realization of θ_t. The realization θ_t is what the principal would like the agent to spend (to be made precise in the utility function below), and in one interpretation represents the spending needed for the agent to fulfill his mission.

Agent strategy space. The agent decides how much to spend in year $t = 1, 2$. The spending strategy is a mapping from the available budget $b_t \in \mathbb{R}_0^+$ (defined in next paragraph) and the realized spending needs into a spending amount, $s_t : [0, b_t] \times \Theta_t \to [0, b_t]$ for $t = 1, 2$. The agent's spending amount s_t is observable to the principal. The agent cannot spend more than his budget b_t for the year, but is able to spend more than is needed (θ_t) for personal benefit (see the agent utility function below). Hence, I call $s_t > \theta_t$ a *misuse of funds*. Consequently, the single action of setting spending amount s_t simultaneously determines to what degree the principal's mission is fulfilled (by fulfilling spending needs) and to what degree the agent acts against the principal's interests.

Budget and fund roll-over. The agent is exogenously granted an annual budget $b \in (0, u]$. This budget is available to the agent every year. In $t = 2$, the available budget b_2 might additionally include unused and rolled-over funds from the previous year, whereas in $t = 1$ only b is available. Thus,

$$b_1 = b, b_2 = b + \Delta(b_1 - s_1),$$

where $\Delta \in [0, 1]$ is the fraction of unused funds from $t = 1$ allowed to be rolled over to the next year. The roll-over rule, represented by Δ, is part of the principal's policymaking, and is set by the principal in $t = 0$.

Auditing. At the end of the year (i.e., once s_t is set by the agent), the principal can audit at a cost of $c_A > 0$ to determine if there was a misuse of funds. The audit technology is perfect and reveals θ_t, and thus also the amount of fund misuse $s_t - \theta_t$. The principal commits to a deterministic audit rule in $t = 0$, one for each year, which maps every possible total spending s_t that year into an audit decision, $A_t : [0, b_t] \to \{0, 1\}$. Hence, the agent is fully aware of the audit rules before moving.

Punishment. In case any fund misuse was detected via audit, an exogenous punishment $p > 2\alpha b$ is inflicted on the agent, which is large enough to deter the agent from misusing. I assume that reducing the agent's utility by p costs the principal $p \cdot c_p$, with $c_p \geq 0$.[7] Punishment can easily be endogenized, but is kept exogenous for simplicity.

Principal policy. This paragraph summarizes the above actions by the principal. The principal acts in $t = 0$, and commits to audit rules A_1, A_2 and a fund roll-over rule $\Delta \in [0, 1]$. The collection $\{A_1, A_2, \Delta\}$ will be called "policy," and the principal's optimal policy is the policy that maximizes expected utility given the agent strategy (best response). Figure 1 plots the timeline of the model. Commitment is important here, because otherwise there is a time-inconsistency problem with costly auditing.

Agent utility function. The agent cares about fulfilling the spending needs of the principal, that is, about doing his job, but also cares about additional spending with personal benefits. As

[6] Because the distributions of the state are known by the principal and the realizations are independent, there is no ratchet effect in this model. That is, the spending amount or realization θ_1 in year 1 does not tell the principal anything new about the realization θ_2 in year 2.

[7] As will become clear later, whether punishment is costly or not does not affect the results. Moreover, results are unchanged if p is endogenized, that is, a choice of the principal.

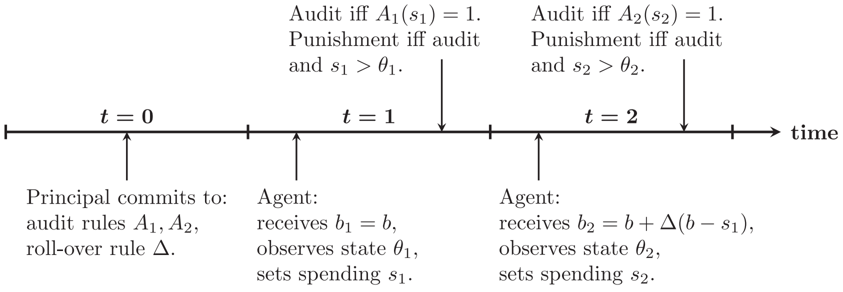


FIGURE 1

TIMING OF PRINCIPAL AND AGENT MOVES, AS WELL AS AUDITS AND PUNISHMENT

described in the introduction, year-end spending is often on items that the principal might not value very highly, but the agent more so. Examples include more luxurious office chairs, new technical gadgets that are fancy but not needed to do the job, or "business trips" to and "conferences" at beach resorts. The agent's utility function is

$$U_{\text{agent}} = \sum_t \left[s_t \cdot \mathbf{1}_{\{s_t \le \theta_t\}} + \mathbf{1}_{\{s_t > \theta_t\}}[\alpha(s_t - \theta_t) + \theta_t]\right] - \sum_t \left[\mathbf{1}_{\{A(s_t)=1, s_t > \theta_t\}} \cdot p\right],$$

with $0 < \alpha < 1$, where $\mathbf{1}_{\{.\}}$ is the indicator function. That is, in the first sum, every dollar spent on fulfilling spending needs (i.e., as long as $s_t \le \theta_t$) yields a marginal utility of 1, whereas every dollar spent above θ_t is fund misuse and yields a marginal utility of $0 < \alpha < 1$. Thus, the agent benefits from additional spending, but not as much as from fulfilling the principal's spending needs.[8] This implies the agent will attempt to misuse funds only once all spending needs are fulfilled. This is consistent with the "use it or lose it" phenomenon of agents observed at the end of the fiscal year, who scramble to spend the remaining funds on something of value to them before they expire, but not at the beginning of the year when fresh funding is available and might crowd out legitimate spending. The utility function also includes the punishment in case of fund misuse and audit (second sum).

Principal utility function. Following Liebman and Mahoney (2017), I model the principal as having a cost of funds $0 < \lambda < 1$. In the framing of a government, this means there is no hard budget constraint, but rather additional funds could be raised via higher taxes or more borrowing, which however comes at marginal cost λ. Similarly, a corporation could change strategy or borrow more at a cost.[9] Because of these costs, the principal only wants spending done with value exceeding λ. Hence, the principal's utility is

$$U_{\text{principal}} = \sum_t s_t \mathbf{1}_{\{s_t \le \theta_t\}} + \theta_t \mathbf{1}_{\{s_t > \theta_t\}} - \sum_t \left[\lambda s_t + \mathbf{1}_{\{A_t(s_t)=1\}}(c_A + \mathbf{1}_{\{s_t > \theta_t\}} \cdot p c_p)\right].$$

Thus, like the agent, the principal receives a marginal utility of 1 from fulfilling spending needs, but unlike the agent, receives no utility from the misuse of funds.[10] Any returned funds are valued at a marginal rate of $0 < \lambda < 1$, or equivalently, the used funds cost λ at the margin. The principal–agent conflict is thus about the additional spending above θ_t, but both agree

[8] While in practice there are undoubtedly agents who value personal spending more than fulfilling the spending needs of their principal, this is not the focus of this article.

[9] The cost of funding can also be interpreted as the opportunity cost of using available funds, for example, funds could be used by another agent yielding a utility of λ to the principal. However, in this article I do not explicitly introduce multiple agents.

[10] Results are similar if the marginal utility of the additional spending to the principal is positive but less than λ, so that the principal does not prefer the additional spending. Hence, setting the marginal utility to zero simplifies the exposition, but is not crucial for the results.

on the spending up to θ_t. Besides the cost of funds, the second sum also includes the costs for auditing and punishment.

The optimal policy is a year 1 audit rule A_1, year 2 audit rule A_2, and roll-over rule Δ (which enters $U_{\text{principal}}$ indirectly by affecting agent spending s_t) which jointly maximize principal expected utility:

$$\max_{A_1, A_2, \Delta} U_{\text{principal}}.$$

3. RESULTS

I proceed by determining the agent fund roll-over reaction to the principal rules, then optimizing the audit rules and roll-over rule separately while holding the other instruments constant. Finally, I combine these results in Subsection 3.6 to obtain the optimal policy.

3.1. *Agent Fund Roll-Over Decision.* Take the principal policy A_1, A_2, Δ as given. Assume the year 2 audit rule has a threshold spending value $\underline{a}_2$, such that there is always auditing above but not weakly below that audit threshold (to be confirmed later), and that $\underline{a}_t \leq b$. Now consider agent year 1 spending s_1 if there is no auditing in year 1. Consider the case $\theta_1 < b$, since roll-over is clearly not optimal otherwise. The marginal utility from spending above θ_1, but below the auditing threshold, is α. The marginal benefit from saving funds in the amount of x, rolling over Δx, and thus setting $b_2 = b + \Delta x$, is the derivative of year 2 expected utility with respect to x:

$$\frac{\partial}{\partial b_2}\left[\int_0^{\underline{a}_2} (\theta_2 + (\underline{a}_2 - \theta_2)\alpha)\mathrm{d}G(\theta_2) + \int_{\underline{a}_2}^{b_2} \theta_2 \mathrm{d}G(\theta_2) + \int_{b_2}^{u} b_2 \mathrm{d}G(\theta_2)\right] \cdot \frac{\partial b_2}{\partial x} = (1 - G(b_2))\Delta.$$

This marginal benefit is clearly nonnegative, less than 1, and strictly decreasing until $G(b_2) = 1$, in which case it is 0. It decreases in b, because the more exogenous budget next year, the less need to roll-over funds to cover the same spending needs. The effect of roll-over policy Δ is ambiguous: It is positive because a larger roll-over share helps fulfill more spending needs for a given savings amount, thus making the transfer "more efficient," and it is negative because the probability of not being able to fulfill spending needs decreases, thus decreasing the need to roll-over. The following expression determines the amount of rolled-over funds x, so that the marginal expected utility from the roll-over equals the marginal utility of misusing funds in year 1:

$$\text{(1)} \qquad \alpha = \Delta(1 - G(b + \Delta x)) \iff x = \hat{x} := \begin{cases} 0, & \text{if} \frac{G^{-1}(1-\frac{\alpha}{\Delta})-b}{\Delta} < 0, \\ b, & \text{if} \frac{G^{-1}(1-\frac{\alpha}{\Delta})-b}{\Delta} > b, \\ \frac{G^{-1}(1-\frac{\alpha}{\Delta})-b}{\Delta}, & \text{else.} \end{cases}$$

That is, $x = \hat{x}$ is the solution to the first-order condition with the boundary condition $\hat{x} \in [0, b]$. Moreover, define $G^{-1}(c) = -\infty$ if $c < 0$, which implies that $\hat{x} = 0$ for any $\Delta < \alpha$.

Hence, even without auditing in year 1, the agent does not misuse all unused funds if parameters and policy are such that $\hat{x} > 0$, because the agent wants to have more funds next year to fulfill additional spending needs.

3.2. *Principal: Audit Rule Classes.*

3.2.1. *Threshold and interval audit rules.* This section proves two results, which characterize the structure of the optimal audit rules. All proofs are provided in the Appendix.

(a) A threshold audit rule: Audit if and only if spending above threshold.

(b) Not a threshold audit rule: There are two disjoint audit intervals. The lower audit region is ineffective, i.e., for any θ_2, the agent spends the same as under the threshold rule above.

NOTES: The line represents the agent budget set and possible spending amounts in year 2. The gray areas are the spending amounts that are audited according to the audit rule.

FIGURE 2

TWO DIFFERENT AUDIT RULES FOR THE SAME SPENDING NEED REALIZATION θ_2

PROPOSITION 1. *If $\hat{x} = 0$, then in both years, there exists an optimal audit rule which is of the threshold type: $A_t(s) = 1$ for any $s > \underline{a}_t$ and $A_t(s') = 0$ for any $s' \leq \underline{a}_t$.*

This is the first main result and tells us we can focus on threshold audit rules for both years if the roll-over rule is not effective in inducing roll-over for any θ_1 (i.e., if $\hat{x} = 0$). This includes the case where roll-over is not allowed. A threshold rule audits all spending amounts above a threshold, and no spending weakly below the threshold. The agent reaction to any audit rule is simple: For a sufficiently small spending need realization θ_t, so that unused funds remain, the agent optimally spends the largest amount that is not audited. Spending more would trigger an audit and punishment, hence is not optimal. Spending less means not misusing as many funds as possible without getting caught.

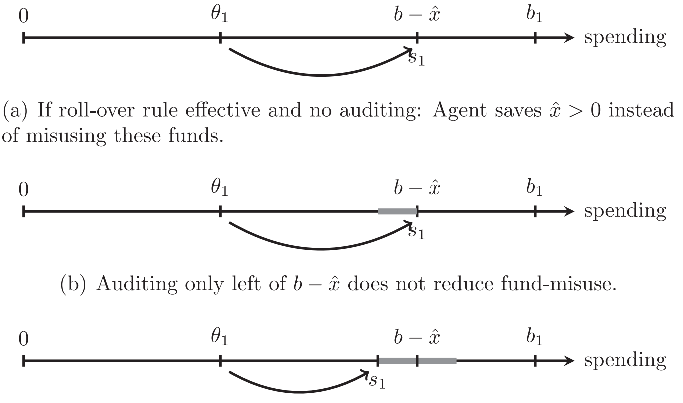
Given this agent reaction function, a threshold rule can replicate the agent spending decision of any nonthreshold audit rule: Take the largest spending amount that is not audited under a nonthreshold rule, then a threshold rule with threshold equal to that amount produces identical outcomes for principal and agent, for any realization θ_t. Figure 2 illustrates these arguments graphically. Hence, any audit rule (including any optimal one) has a corresponding threshold audit rule that is outcome-equivalent. Therefore, we can focus on this simple class of audit rules to determine the optimal policy.

Next, in year 1 and if agents want to roll-over funds without auditing (i.e., if $\hat{x} > 0$), then the optimal audit rule is an interval rule that takes a slightly different shape. The year 2 audit rule remains a threshold rule. This is the second main result, which implies that the optimal year 1 audit rule changes its structure once fund roll-over is allowed and effective. Consequently, introducing fund roll-over has important implications for auditors.

PROPOSITION 2. *If $\hat{x} > 0$, then in year 1, there exists an optimal audit rule which is of the interval type: $A_1(s_1) = 1$ for any $s_1 \in (\underline{a}_1, \overline{a}_1)$, and $A_1(s_1') = 0$ for any $s_1' \notin (\underline{a}_1, \overline{a}_1)$, with $(b - \hat{x}) \in [\underline{a}_1, \overline{a}_1]$. For any $\underline{a}_1 \leq b - \hat{x}$, the upper border of the interval is given by*

$$\overline{a}_1(\underline{a}_1) = \frac{2b\Delta(1+\Delta) + 2u(\alpha - \Delta) - \underline{a}_1\Delta^2}{\Delta^2}. \tag{2}$$

This is the second main result and tells us that, to find the optimal audit rule, we can focus on the class of interval audit rules in the first year if $\hat{x} > 0$. With an interval audit rule, spending in the interval $s_1 \in (\underline{a}_1, \overline{a}_1)$ is audited, whereas spending outside of the interval is not. The optimal interval border $\underline{a}_1$ is discussed in the online appendix, with a summary in Subsection 3.5.



(a) If roll-over rule effective and no auditing: Agent saves $\hat{x} > 0$ instead of misusing these funds.

(b) Auditing only left of $b - \hat{x}$ does not reduce fund-misuse.

(c) Interval rule: Auditing on both sides of $b - \hat{x}$ can reduce fund-misuse.

NOTES: Reducing fund misuse further requires an audit rule to audit on both sides of $s_1 = b - \hat{x}$ as in (c).

FIGURE 3

IF THE FUND ROLL-OVER RULE IS EFFECTIVE, THEN THE AGENT SPENDS $s_1 = b - \hat{x}$ ABSENT AUDITING TO SAVE FUNDS FOR NEXT YEAR (A)

The argument why an interval audit rule is optimal is as follows: If the roll-over rule is effective, then the agent does not spend all funds in year 1 even if there is no auditing, provided the realization θ_1 is small enough so that there are funds left over. The agent prefers to save an amount $\hat{x} > 0$ to roll-over the amount $\Delta\hat{x} > 0$ and to be able to fulfill more spending needs next year (see the construction of $\hat{x}$ above). Hence, the amount $\hat{x}$ is saved instead of misused. This is graphically shown in Figure 3(a). In this case, the roll-over rule prevents some fund misuse without audit cost.

If the principal wants to reduce fund misuse further, then she cannot audit only spending amounts left of $b - \hat{x}$ to push spending down. This is because the agent would then prefer to spend either $b - \hat{x}$ or just above, and hence fund misuse would not decrease (Figure 3b).

To reduce fund misuse further, the principal has to audit on both sides of $b - \hat{x}$. Auditing to the left of this amount pushes spending down, whereas auditing to the right of $b - \hat{x}$ makes the lower spending amount incentive-compatible, because it prevents the agent from spending more than $b - \hat{x}$. See Figure 3(c) for an illustration. The optimal interval rule chooses the upper border of the audit interval, $\overline{a}_1$, as a function of the lower border $\underline{a}_1$ to make agents with small θ_1 just indifferent between spending $s_1 = \underline{a}_1$ (below the interval) and spending $s_1 = \overline{a}_1$ (above the interval, implying more fund misuse).

An interval audit rule might not audit the very highest spending amounts close to $b_1 = b$, unlike a threshold rule. This is because the agent prefers saving some funds to spending everything if the spending need θ_1 is small enough. Consequently, auditing very high spending amounts is not necessary in these cases, and the principal saves audit costs because the fund roll-over rule is effective. In other words, the principal could use a threshold rule even if $\hat{x} > 0$, but this is not optimal, because she could audit less without increasing fund misuse. A special case of the interval rule is $\underline{a}_1 = \overline{a}_1 = b - \hat{x}$, which does not audit at all.[11]

To focus the analysis of the optimal policy, I will restrict attention to the cases $\underline{a}_t \leq b$. Allowing $\underline{a}_t > b$ adds a lot of additional case distinctions to the analysis with comparatively little

[11] The logic why an interval rule is optimal would apply also in case of probabilistic audits. If $\hat{x} > 0$, observing the highest spending amounts means there is no fund misuse. The agent spends only so much to fulfill spending needs, never to misuse funds. Thus, auditing these highest spending amounts is a waste, and an optimal audit rule puts zero audit probability on these amounts, independent of whether auditing may be deterministic or probabilistic.

additional insight.[12] This constraint is without loss of generality for $\underline{a}_1$, since $\underline{a}_1 = b$ implies no auditing, and $\underline{a}_1 = 0$ implies always auditing, and all cases in between are covered. However, if $b_2 > b$ and $u > b$, then this constraint might bind for $\underline{a}_2$.

ASSUMPTION. *The exogenous upper bound of the audit threshold is b, that is, $\underline{a}_t \leq b$.*

3.2.2. *Which organizations should use an interval audit rule?* Based on Proposition 2, we can determine what kind of organizations should use interval audit rules instead of threshold audit rules. It shows that organizations should use interval audit rules if they allow fund roll-over and where the fund roll-over rule is effective in inducing agents to roll-over funds for sufficiently small spending need realizations (i.e., where $\hat{x} > 0$).

The comparative statics for $\hat{x}$ (see (1)) show that these tend to be organizations with, first, smaller agent utility from fund misuse α. Since α is the opportunity cost of saving and rolling over, a lower α clearly encourages more fund roll-over. A smaller α in practice might mean less spending discretion for the agent or that fund misuse is harder due to additional checks before spending is approved.

Second, it tends to be organizations with smaller budget b relative to maximum spending needs u, that is, agents that are more underfunded. In practice, these could be organizations or divisions that get a lower budget than they asked for or get a budget that is below an amount they need in case of unfortunate contingencies. These agents want to save more for “a rainy day” if they can, to help with the underfunding next year, and this allows the principal to audit less at the top of the budget set via an interval rule.

The effect of the audit rule Δ is ambiguous on $\hat{x}$. On the one hand, allowing a larger share of unspent funds to be rolled over makes saving and fund roll-over more attractive compared to fund misuse, as more of the saved funds are available for use next year. On the other hand, a larger Δ means fewer funds need to be saved to have a certain amount next year, thus making roll-over less attractive.

3.3. *The Optimal Roll-Over Rule.* Let us first determine the roll-over share that maximizes the agent saving $\hat{x}$ in year 1 (for a sufficiently small realization θ_1). Plugging the inverse of the uniform CDF into the expression for the agent saving amount $\hat{x}$ in (1) yields

$$\left(\frac{1}{\Delta} - \frac{\alpha}{\Delta^2}\right)u - \frac{b}{\Delta}.$$

Differentiating with respect to Δ once and setting to zero yields the necessary first-order condition

$$-\frac{u}{\Delta^2} + \frac{2\alpha u}{\Delta^3} + \frac{b}{\Delta^2} = 0.$$

The roll-over share which maximizes agent saving is the solution to this first-order condition subject to the upper bound:[13]

$$\Delta_{\max\hat{x}} := \min\left\{\frac{2\alpha u}{u-b}, 1\right\} > \alpha > 0. \tag{3}$$

[12] For example, without this restriction, the agent marginal utility of rolling-over funds discontinuously drops as the rolled-over amount increases so that $b_2 \to \underline{a}_2$, which requires additional case distinctions for interior and corner solutions for the optimal roll-over amount.

[13] While the optimization problem is not concave, the proof to Proposition 3 in the Appendix shows that $\hat{x}$ is increasing below and decreasing above $\Delta = \Delta_{\max\hat{x}}$, making it a maximum in $\Delta \in [0, 1]$.

Proposition 3 is the third main result. It takes an $\underline{a}_1$ as given. It shows that either a potentially interior $\Delta^* \in (\Delta_{\max\hat{x}}, 1]$ or the maximum $\Delta = 1$ is optimal.[14] Consequently, since $\Delta_{\max\hat{x}} > \alpha > 0$, allowing at least some roll-over of unused funds is weakly better than no roll-over.

PROPOSITION 3 (OPTIMAL ROLL-OVER RULE). *Assume the principal uses an interval audit rule if* $\hat{x} > 0$ *and a threshold audit rule if* $\hat{x} = 0$. *Fix an* $\underline{a}_1$.

(i) *The optimal budget roll-over rule is a* $\Delta^* \in (\Delta_{\max\hat{x}}, 1]$ *if all of the following conditions are fulfilled:*
   - $\hat{x} > 0$ *at* $\Delta = \Delta_{\max\hat{x}}$, *and*
   - $\bar{a}_1(\underline{a}_1) < b$ *at* $\Delta = \Delta_{\max\hat{x}}$, *and*
   - $\Delta_{\max\hat{x}} < 1$, *and*
   - $\hat{x} < b$ *at* $\Delta = 1$.

   If, in addition, λ *is sufficiently large but less than 1, then a* $\Delta^* \in (\Delta_{\max\hat{x}}, 1)$ *is optimal. That is, positive but partial roll-over is optimal.*

(ii) *A full roll-over policy* ($\Delta^* = 1$) *is optimal if any of the four conditions in i. fail, that is, if*
   - $\hat{x} = 0$ *at* $\Delta = \Delta_{\max\hat{x}} < 1$, *or*
   - $\bar{a}_1(\underline{a}_1) > b$ *at* $\Delta = \Delta_{\max\hat{x}}$, *or*
   - $\Delta_{\max\hat{x}} = 1$, *or*
   - $\hat{x} = b$ *at* $\Delta = 1$.

Subsection 3.3.1 explains the intuition for why partial roll-over can be optimal, and Subsection 3.3.2 explains the proposition and its conditions.

3.3.1. *Why partial roll-over of unused funds can be optimal.* To understand why sometimes $\Delta = 1$ is optimal and sometimes $\alpha < \Delta < 1$, we have to distinguish two purposes of the roll-over rule. First, given funds are saved in year 1—either due to the threat of auditing or because the agent wants to roll-over funds—it is weakly better for agent and principal alike to roll-over as much as possible. This is because more roll-over increases the chances to fulfill additional spending needs in year 2.[15]

Second, the roll-over rule also determines $\hat{x}$, how much the agent wants to save instead of misuse in year 1. Hence, the possibility of rolling-over funds induces the agent to misuse fewer funds, which also benefits the principal. The amount saved by the agent $\hat{x}$ can be locally decreasing, so that there is a unique $\Delta_{\max\hat{x}} < 1$ maximizing the saved amount $\hat{x}$. Figure 4 plots an example. Consequently, in these cases, the principal has to trade-off the maximum $\Delta = 1$ (first purpose) with $\Delta = \Delta_{\max\hat{x}} < 1$ (second purpose).

The intuition why $\Delta_{\max\hat{x}}$ can be less than 1 is as follows: Whereas the agent wants to misuse leftover funds in year 1 to get benefit α, he also wants to roll-over funds to be able to fulfill more spending needs next year (if budget b is not too large). This is because without the roll-over, there is a larger probability that not all spending needs can be met, which reduces agent utility. The roll-over rule Δ has two different effects on the agent benefit of rolling-over funds. On the one hand, Δ makes the budget transfer from one year to the next more efficient (fewer funds are returned to the principal), which increases the benefit of rolling over. On the other hand, a larger Δ increases the available budget next year for a given savings amount, which decreases the probability of not being able to fulfill all spending needs next year, thus

[14] Since the optimization problem is not well-behaved, it is difficult to determine the exact solution in the interval $\Delta \in (\Delta_{\max\hat{x}}, 1]$ in part (i). That is, a corner solution may be better than an interior solution, and interior solutions are not guaranteed to be unique. Hence, the result in (i) is given as an interval, whereas case (ii) is guaranteed to be $\Delta = 1$. For large enough λ, (i) establishes that the optimal Δ must be in the interior of the interval.

[15] And, given $\underline{a}_2 \leq b$, any rolled-over funds, which lead to $b_2 > b$, cannot be misused in year 2, since that would trigger an audit.

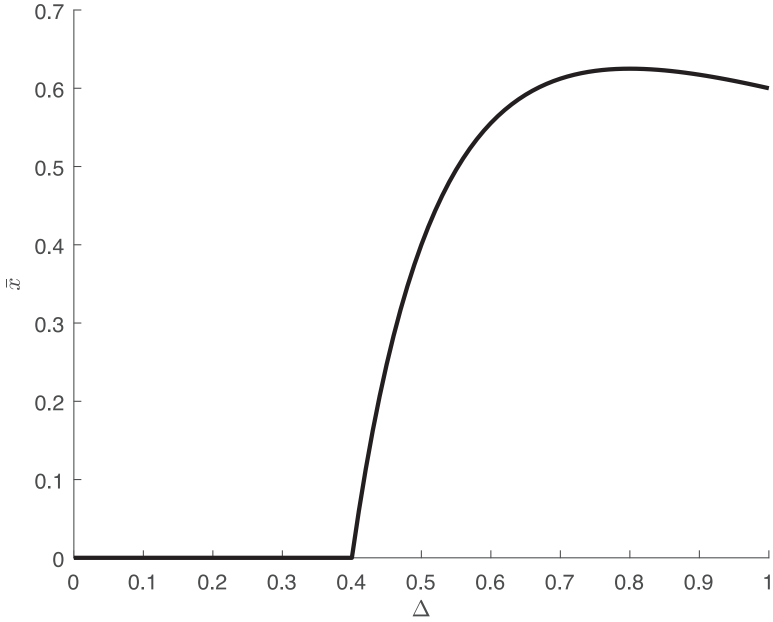


FIGURE 4

PLOT OF AGENT SAVING $\hat{x}$ DEPENDING ON Δ

NOTES: PARAMETER VALUES: $u = 2, b = 1, \alpha = 0.2$. THE MAXIMUM IS AT $\Delta_{\max\hat{x}} = 0.8$. THUS, THE OPTIMAL Δ MUST EXCEED 0.8

reducing the benefit of rolling over more. In other words, a larger roll-over share Δ makes less roll-over necessary to insure against a binding budget constraint in the future, so the agent misuses instead of saves more funds. This hurts principal utility. Hence, it can be optimal to "tax" the roll-over ($\Delta < 1$) to induce more saving and less misuse by the agent. An analogy from another context is retirement savings, where the optimal savings amount can increase if investment returns decrease.

3.3.2. *Explaining the optimal roll-over rule.* Proposition 3 part (i) is the case where partial roll-over ($\Delta < 1$) can be optimal. It requires that $\Delta_{\max\hat{x}} < 1$, otherwise there is no trade-off between inducing more saving and rolling over more funds. $\Delta_{\max\hat{x}} < 1$ requires a sufficiently small agent utility from fund misuse, α, because that is the opportunity cost of rolling over, and a sufficiently small budget b. The smaller budget is needed, so there is a larger chance the agent cannot meet his spending needs next year, hence he is more willing to roll-over even for lower values of Δ. This is not a precautionary savings motive, as the agent is risk-neutral, but an expected value calculation that rolling over is more attractive if the budget is small.

The trade-off between $\Delta < 1$ and $\Delta = 1$, as discussed above, consists of preventing more fund misuse (at a marginal utility of λ) and saving audit costs, or of fulfilling more spending needs (at a net marginal utility of $1 - \lambda$). Consequently, if λ is sufficiently large, then $\Delta < 1$ is optimal, since fulfilling additional spending needs next year is not as important as preventing fund misuse and saving audit costs this year.

$\Delta < 1$ can be optimal only if b and α are not too large, so that the agent wants to save funds at all, that is, only if $\hat{x} > 0$ for $\Delta_{\max\hat{x}} < 1$. But b and α also cannot be too small, since otherwise the agent wants to save all unused funds even if $\Delta = 1$, which would then be optimal. Hence, $\Delta^* < 1$ also requires $\hat{x} < b$ at $\Delta = 1$.

In case (ii), which captures all cases where $\Delta_{\max\hat{x}} = 1$ maximizes the amount saved by the agent, the maximum $\Delta^* = 1$ is optimal. This is because there is no conflict between inducing the maximum saving by the agent and rolling-over everything that is saved. Moreover, if

$\Delta_{\max\hat{x}} < 1$ but no saving by the agent can be induced for any Δ—for example, if b is large—then $\Delta^* = 1$ is also optimal.

Informally, for partial roll-over to be optimal, it requires a medium level of b and a sufficiently small benefit of fund misuse α, so that partial roll-over increases the savings amount by the agent. In addition, the cost of funds λ needs to be sufficiently large, so that the principal favors preventing fund misuse in year 1 via more agent saving over fulfilling more spending needs in year 2 via more roll-over. Otherwise, full roll-over is optimal.

3.3.3. *Which organizations should have stricter roll-over?* Based on Proposition 3, we can determine what kind of organizations should use stricter roll-over, in the sense that the optimal roll-over share Δ^* is smaller.[16] First, a larger cost of funds λ weakly decreases Δ^*. As explained in the previous section, this is because getting the agent to save and roll-over is more important than fulfilling additional spending needs in year 2 if λ is larger. Hence, organizations that have a harder time borrowing should be less generous in their roll-over.

Second, a smaller agent utility from fund misuse α weakly decreases Δ^*. As fund misuse becomes less attractive (α decreases), the principal can induce more agent saving by increasing the "tax" on the roll-over, which reduces fund misuse in year 1. A smaller α in practice might be less spending discretion or that fund misuse gets harder due to additional checks before spending is approved.

The effect of budget b is ambiguous. A smaller budget b tends to increase the savings amount and hence the principal can get away with taxing the roll-over more, as with a decrease in α discussed earlier. This favors a smaller Δ^*. But at the same time, a smaller b makes fulfilling spending needs in year 2 harder, so the principal would like to allow more roll-over to counter a more-likely-to-bind budget constraint. In the extreme case, if b is sufficiently small, then the agent is so underfunded that he wants to save everything independent of Δ, thus $\Delta^* = 1$ is optimal. For the same reasons, the effect of a change in the maximum spending need u is ambiguous.

There is a relationship between partial roll-over and interval audit rules: A partial roll-over is optimal only if it increases the agent savings amount $\hat{x} > 0$ (hence, if there is no saving independent of the roll-over rule, then full roll-over is optimal). And, according to Proposition 2, interval auditing is also optimal if and only if the agent is saving leftover funds.

What is the minimum level of Δ in practice? For the agent to want to save and roll-over any funds, the share of unused funds that is rolled-over must exceed α, which is the marginal value the agent receives from misusing funds. Hence, in this model with piecewise linear utility functions, if agents at the margin get 50% of utility from misusing funds compared to fulfilling spending needs, then *more than* 50% of unused funds must be rolled over to incentivize agents not to misuse funds. Thus, the minimum Δ is bounded below by the benefit of fund misuse.

3.3.4. *Comparison to budget rules in practice.* As the literature section shows, most public-sector organizations in the United States do not allow agents to roll-over unused funds (e.g., Jones, 2005; McPherson, 2007; Liebman and Mahoney, 2017), which conflicts with the finding here where at least some roll-over is weakly better. But there are a few documented cases where fund roll-over was allowed in practice.

The UK governmental roll-over rule, introduced in 2010, allows for a full roll-over of unused funds ($\Delta = 1$), but only for up to 0.75% (large agencies) or up to 4% (small agencies) of the budget (HM Treasury, 2021, p. 17). In the cases where my model finds $\Delta^* < 1$ to be optimal, the UK implementation might be imperfect in two respects: First, $\Delta = 1$ does not discourage as much fund misuse as $\Delta^* < 1$ as explained earlier, and second, the 4% upper bound

[16] The following comparative statics are weak inequalities, in the sense that they change an interior solution in the stated direction, but if a corner solution $\Delta^* = 1$ is optimal, then a small change in parameters will not change the optimal roll-over share.

means the UK rule can at most prevent fund misuse in the amount of 4% of the budget. A larger percentage could potentially prevent more fund misuse.

There is at least one case where a (local) governmental body allowed its agencies to roll-over some, but not all, of their unused funds to the next year. The State of Washington's Saving Incentive Program from 1997 allowed agencies to roll-over 50% of unused funds for next year. Interestingly, the 50% were set not because that number is optimal in reducing fund misuse, as in this model. Instead, the other 50% were promised to the education sector, which made this reform politically feasible (e.g., Jones, 2005, p. 152; Miller et al., 2007).

3.4. *The Optimal Year 2 Audit Rule.* Proposition 4 derives the optimal audit threshold in year 2. The year 2 audit rule can condition on b_2, and is therefore conditionally independent of the year 1 audit rule and the roll-over rule.

PROPOSITION 4 (OPTIMAL YEAR 2 AUDIT RULE). *As shown in Proposition 1, the optimal audit rule can be represented as a threshold rule, with $A_2(s_2) = 1$ if and only if $s_2 > \underline{a}_2$, and $A_2(s_2) = 0$ otherwise.*

(i) The optimal audit threshold for year 2 is $\underline{a}_2^* = \min\{\frac{c_A}{\lambda}, b\}$ if
- $b = u$, *or*
- $b < b_2$, *or*
- $b = b_2 < u$, $c_A/\lambda \leq b$ *and* $\frac{b_2^2\lambda}{2} + \frac{c_A^2}{2\lambda} \geq uc_A$.

(ii) The optimal threshold is $\underline{a}_2^* = b$ if $b = b_2 < u$, $c_A/\lambda \leq b$ and $\frac{b_2^2\lambda}{2} + \frac{c_A^2}{2\lambda} < uc_A$. That is, in this case, there is no auditing even if all budget is spent.

Proposition 4 shows that the optimal audit rule, a threshold rule, is simple in that either the optimal threshold equals the interior solution $\underline{a}_2^* = c_A/\lambda$ or the corner solution $\underline{a}_2^* = b$.

When setting the audit threshold, a larger threshold implies less auditing, because an audit is triggered only for spending amounts exceeding the threshold. Agents with a spending need realization exceeding the threshold ($\theta_2 > \underline{a}_2$) do not misuse funds to avoid punishment. Agents with low spending needs below the threshold ($\theta_2 < \underline{a}_2$) misuse funds to stay just below the threshold. Hence, a larger threshold implies more fund misuse by the agent, who can get away with misusing more without being audited. But the larger audit threshold also saves audit costs. This is the key trade-off when setting the audit threshold. The interior solution $\underline{a}_2^* = c_A/\lambda$ optimally trades off the additional fund misuse from increasing the threshold and the saved audit cost. Consequently, a larger audit cost c_A or a smaller disutility from additional fund misuse (equal to the cost of funds λ) increases the audit threshold, that is, leads to less auditing in the interior solution.

If $b = u$—where the yearly budget is enough to fulfill all spending needs for certain—or if $b < b_2$—where some funds were rolled-over from the previous year—then the interior solution is optimal unless $c_A/\lambda \leq b$ is binding, in which case the corner solution is optimal. These are the technically well-behaved cases, because there are no discontinuities in the principal expected utility function for $\underline{a}_2 \in [0, b]$.

However, if $b = b_2 < u$, then there can be a discontinuity at $\underline{a}_2 = b$. To understand why there is a discontinuity, note that $b_2 < u$ implies that the budget constraint for the agent is binding for all realizations $\theta_2 \in (b_2, u]$. Consequently, a probability mass of at least $1 - G(b_2)$ is spending $s_2 = b_2$. So when comparing the principal utility at $\underline{a}_2 = b_2 - \varepsilon$ with $\varepsilon > 0$ small (which audits $s_2 = b$) and at $\underline{a}_2 = b_2$ (which does not audit $s_2 = b$), this probability mass introduces a discontinuous jump in saved audit costs. Case (ii) in Proposition 4 shows that the corner solution might be optimal due to the discontinuity even if the interior solution is in fact in the interior. This happens in particular if b_2 is sufficiently small, whereas the interior is favored if λ is sufficiently large.

Budget b can also have an effect on the optimal audit rule. Whereas the audit threshold of the interior solution does not depend on the budget, it does in part determine whether a corner solution with little auditing is optimal. As $\underline{a}_2^* = \min\{c_A/\lambda, b\}$ in part (i) suggests, a larger budget tends to favor the interior solution and increase audit activity. For small budgets, the corner solution prescribes less auditing, since there is less room for fund misuse. For large budgets, the probability of the budget exceeding spending needs increases, so fund misuse is more problematic and auditing more useful.

3.5. *The Optimal Year 1 Audit Rule.* Closed-form solutions for the optimal year 1 audit threshold (if the roll-over rule is ineffective, $\hat{x} = 0$) and for the optimal year 1 audit interval (if the roll-over rule is effective, $\hat{x} > 0$) are derived in the online appendix. The expressions are long and complex, so I only summarize the main take-aways here. The main qualitative differences between threshold and interval rules are discussed with Propositions 1 and 2.

First, Proposition A.6 in the online appendix shows that the optimal audit rule in year 1 if $\hat{x} = 0$ is more complex than the one in year 2, because the principal has to take the effects on year 2 into account when setting audit threshold $\underline{a}_1$. Setting a lower audit threshold—that is, auditing more—means there is less fund misuse and more saving by the agent. If fund roll-over is allowed, then more auditing implies a larger budget in year 2 for small θ_1-realizations, which can benefit both agent and principal by fulfilling additional spending needs. Consequently, the principal in year 1 tends to audit more than in year 2 in order to induce more roll-over (Prop. A.6(iii)).

Otherwise, the year 1 audit threshold behaves very much like the year 2 audit threshold, increasing with audit costs and decreasing with the cost of funds. For large b, the year 1 and year 2 audit thresholds are identical. For the maximum budget $b = u$, this is easy to see, because then both years are ex ante identical, since the second year has enough budget to fulfill all spending needs even without fund roll-over.

Second, Proposition A.7 in the online appendix shows that the optimal interval audit rule (if $\hat{x} > 0$) in some situations audits less and prevents less fund misuse than if the principal was forced to use a threshold rule. This is because marginally preventing more fund misuse under an interval requires both a reduction of the lower audit interval border $\underline{a}_1$ *and* an increase in the upper audit interval border $\overline{a}_1$ to keep spending $s_1 = \underline{a}_1$ incentive-compatible. Consequently, the marginal cost of preventing more fund misuse is higher under an interval rule than under a threshold rule, where the audit region is only extended in one direction. Thus, the interval rule audits less and prevents less fund misuse in some situations. But the principal is nevertheless weakly better off with the interval rule, because the interval audit rule allows her to save audit costs.[17]

Unless the optimum is a corner solution, the optimal audit interval expands with a lower audit cost and a larger cost of funds. For very large audit costs, the audit interval is empty, so only the roll-over incentive prevents maximum fund misuse by the agent. For very small audit costs, the optimal audit interval can be large enough to include spending $s_1 = b$, and is thus outcome-equivalent to a threshold rule.

3.6. *Optimal Policy.* The previous propositions maximized principal expected utility for each instrument, holding the others constant. However, the optimal policy requires all instruments to be jointly optimal. The optimal policy needs to specify the following instruments. First, the audit rule in year 1, which needs to specify an audit threshold $\underline{a}_1$ and whether the audit rule is a threshold or an interval rule. Second, the audit rule in year 2, which is always a threshold rule, and needs to specify the spending threshold $\underline{a}_2$ above which the agent is audited. Third, the roll-over rule Δ, which is the share of saved funds in year 1 that is added to the agent's budget in year 2.

The optimal year 2 audit threshold is independent of the other instruments, so the proposition applies directly. However, the roll-over rule and year 1 audit rule are interdependent.

[17] Notice the difference between marginal and total expected audit cost: An interval rule has a higher marginal cost of auditing compared to a threshold rule. But, because the interval rule prevents some fund misuse for free, due to the agent roll-over $\hat{x} > 0$, it has a lower total expected audit cost.

According to Proposition 3, the optimal roll-over rule is $\Delta^* = 1$ if $\alpha \geq 1/2$ or if b is sufficiently large, independent of $\underline{a}_1$. But if $\Delta^* \in (\Delta_{\max\hat{x}}, 1]$, then the exact optimum in that interval can depend on $\underline{a}_1$.

In the following, "no auditing" in year 1 means not auditing any possible spending amount. "No auditing" in year 2 means $\underline{a}_2 = b$ in year 2, which is the least auditing possible under the assumption of $\underline{a}_t \leq b$,[18] so no spending amount $s_2 \in [0, b]$ is audited. "Active auditing" means at least one spending amount in $s_t \in [0, b]$ is audited.

Overall, in the optimal policy, full fund roll-over can be consistent with auditing and with no auditing. Similarly, auditing can be consistent with full and only partial fund roll-over, depending on parameters. However, if the year 1 audit rule is a threshold rule, then only full fund roll-over can be optimal, since the threshold rule does not benefit from more agent saving. An interval rule in year 1, on the other hand, can be consistent with both full and partial fund roll-over. Proposition 5 describes the optimal policy in several distinct situations, and follows from the previous propositions.[19] Online appendix B.2 derives the optimal policy even for ambiguous cases numerically.

PROPOSITION 5.

(i) *If $\alpha > 1/2$, then full fund roll-over is optimal ($\Delta^* = 1$) independent of the audit rules.*

(ii) *Fix u and α. If b is sufficiently large and c_A/λ is sufficiently small relative to b, then the optimal policy is active auditing in both years ($\underline{a}_1^t = \underline{a}_2^* = c_A/\lambda < b$) and a full roll-over of unused funds ($\Delta^* = 1$). The audit rules are threshold rules.*

(iii) *Fix u. If α is sufficiently small, b is sufficiently small but $b > \frac{(1-\alpha)u}{2}$, and c_A/λ is sufficiently small relative to b, then the optimal policy is active auditing in both years, with an interval rule in year 1 ($\underline{a}_1^i \leq 2c_A/\lambda$) and a threshold rule in year 2 ($\underline{a}_2^* = c_A/\lambda$). A partial roll-over of unused funds may be optimal ($\Delta^* \in (\Delta_{\max\hat{x}}, 1]$).*

(iv) *Fix u and $b \leq \frac{(1-\alpha)u}{2}$. Then the optimal policy is an interval audit rule with no auditing ($\underline{a}_1^i = b - \hat{x}$) in year 1 and a full roll-over of unused funds ($\Delta^* = 1$). In year 2, auditing uses a threshold rule, either with active auditing if c_A/λ is sufficiently small relative to b, or with no auditing if c_A/λ is sufficiently large relative to b.*

(v) *Fix u. If α is sufficiently small, b is sufficiently small but $b > \frac{(1-\alpha)u}{2}$, and c_A/λ is sufficiently large relative to b, then the optimal policy is no auditing in both years ($\underline{a}_1^* = \underline{a}_2^* = b$). A partial roll-over of unused funds may be optimal ($\Delta^* \in (\Delta_{\max\hat{x}}, 1]$).*

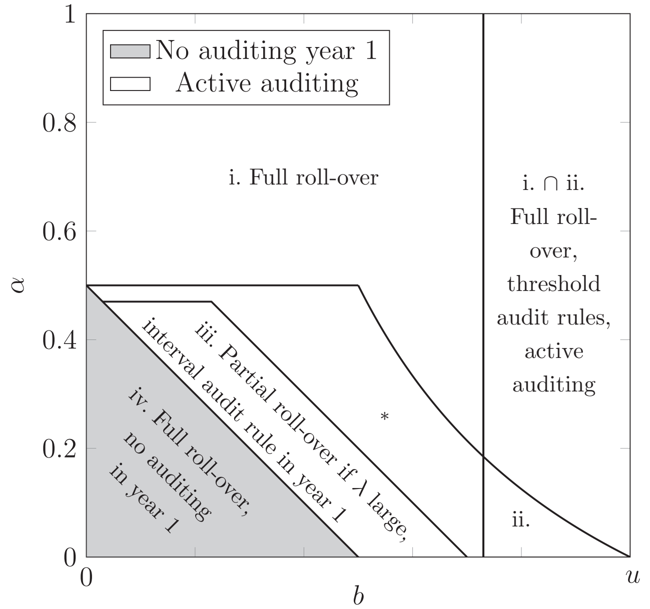
Intuitive explanations for these results are as follows: In part (i), if the agent benefit from fund misuse is large enough, then no partial roll-over ($\Delta < 1$) can induce the agent to save more (compared to full roll-over $\Delta = 1$), as fund misuse is too attractive in comparison. Since the only reason to use partial roll-over is that this induces more saving by the agent, this condition implies that full roll-over is optimal.

In part (ii), if audit costs in relation to the cost of funds (c_A/λ) are sufficiently small and budget b is sufficiently large, then threshold audit rules with active auditing (i.e., low audit thresholds) are optimal. This is because lower audit costs and a higher cost of funds (which is the same as the cost of fund misuse) make auditing attractive to prevent more fund misuse. Moreover, since budget b is large, it is not possible to induce the agent to save more via partial roll-over (similar in part (i)). Thus, full roll-over ($\Delta = 1$) is optimal.

In part (iii), a small agent benefit to fund misuse α means that fund misuse is not attractive and a small budget b means it is likely that not all spending needs can be met next year without roll-over. Thus, the agent would rather save to roll-over than misuse all leftover funds, and

[18] If funds are rolled over by the agent and are at least partially used, then spending is $s_2 > b = \underline{a}_2$ and hence triggers an audit. Otherwise no audit occurs.

[19] The conditions "c_A/λ sufficiently small" and "α sufficiently small" mean these parameters should be sufficiently close to zero. "b sufficiently large" means it should be sufficiently close to u, the maximum realization of θ_t and upper bound for b.



NOTES: The Roman numerals correspond to the cases in Proposition 5. Area * represents indeterminate cases, where either full or partial roll-over may be optimal.

FIGURE 5

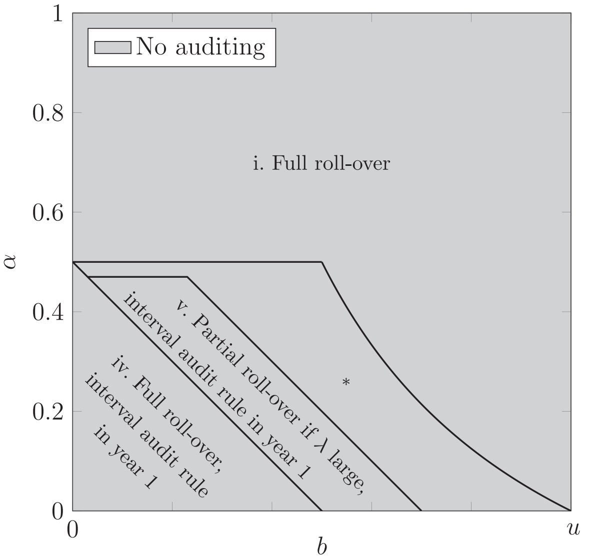
SCHEMATIC PLOT OF OPTIMAL POLICY AS A FUNCTION OF AGENT BENEFIT FROM FUND MISUSE α AND BUDGET b, WHEN c_A/λ IS SUFFICIENTLY SMALL

so observing spending in year 1 close to b means that the spending need θ_1 requires such large spending. Hence, auditing spending close to b is unnecessary. This makes an interval audit rule optimal, with a large audit interval since audit costs relative to the cost of funds (c_A/λ) are small. In year 2, a threshold rule with active auditing is optimal (similar in part (ii)). The constraint $b > \frac{(1-\alpha)u}{2}$ means that the agent does not want to save all possible leftover funds with a full roll-over rule, hence a partial roll-over ($\Delta < 1$) can increase the savings amount further, which may be optimal, in particular if λ is large (see Proposition 3).

In part (iv), $b \leq \frac{(1-\alpha)u}{2}$ means the budget is so low that the agent wants to save all unused funds for any possible θ_1 realization even if full roll-over is allowed. Hence, there is no fund misuse in year 1. And, thus, there is no reason to use partial roll-over to increase saving further, so a full roll-over is optimal. In year 2, depending on the audit costs relative to the cost of funds, either active or no auditing is optimal.

Finally, in part (v), high audit costs relative to the cost of funds make no auditing optimal. The small agent benefit of fund misuse and small budget mean that the agent wants to save leftover funds for roll-over if possible. A partial roll-over may be optimal due to $b > \frac{(1-\alpha)u}{2}$ as in part (iii), as it can increase the agent savings amount further compared to a full roll-over rule.

Figure 5 schematically plots Proposition 5 for the case where c_A/λ is sufficiently small. The lower left triangle is determined by the condition $b \leq \frac{(1-\alpha)u}{2}$, where auditing in year 1 is unnecessary, which is illustrated via the gray shading. Outside the gray area, active auditing is optimal due to audit costs being small. The region just above the gray triangle, with $b > \frac{(1-\alpha)u}{2}$ but α and b still small, is the area where partial roll-over may be optimal. The area just above it, marked by *, is indeterminate and not covered in Proposition 5: depending on parameters, either partial or full roll-over is optimal. In the top and right areas with large α and large b, full roll-over with active auditing are optimal.



NOTES: The Roman numerals correspond to the cases in Proposition 1. Area * represents indeterminate cases, where either full or partial roll-over may be optimal.

FIGURE 6

SCHEMATIC PLOT OF OPTIMAL POLICY AS A FUNCTION OF AGENT BENEFIT FROM FUND MISUSE α AND BUDGET b, WHEN c_A/λ IS SUFFICIENTLY LARGE

Figure 6 schematically plots Proposition 5 for the cases where c_A/λ is sufficiently large. In these cases, no auditing is optimal. As before, partial roll-over may be optimal in the area just above $b = \frac{(1-\alpha)u}{2}$ where α and b is small, otherwise a full roll-over policy is optimal.

4. EXTENSIONS AND ROBUSTNESS

Extensions and the discussion of robustness are delegated to the online appendix. I will only briefly summarize the take-aways here.

Online appendix B.1 lets the principal decide the budget amount b as well as the audit policies and roll-over rule as before. Then it can be analytically shown that the principal underfunds ($b^* < u$) the agent if audits (c_A) and funds (λ) are very costly. Thus, underfunding the agent to prevent misuse is the optimal policy. In these cases, partial roll-over as well as interval auditing can be optimal as well. This means the principal knows that there is a positive probability the agent cannot fulfill all spending needs, reducing the principal utility, but this is better than the expected fund waste with a higher budget. Otherwise, either with sufficiently small audit costs or sufficiently small cost of funds, the maximum budget is optimal.

Online appendix B.2 calculates the optimal policy (budget, roll-over policy, audit rules, and punishment) for various parameter profiles, and demonstrates when partial roll-over and interval audit rules are optimal. Online appendix B.3 argues that, if endogenous budgets can be different in both years, then the year 2 budget should be weakly smaller than the year 1 budget. In practice, this means budgets could be reduced over time, as the possibility to save up over time offsets the lower budget, and indeed incentivizes the agent to save more. Online appendix B.4 discusses how probabilistic auditing could be easily allowed if punishment was endogenous. Allowing probabilistic auditing produces qualitatively similar results as in the present model, except it effectively reduces audit costs.

Online appendix B.5 discusses the effect of using a distribution other than the uniform distribution for the spending needs realizations θ_t. In short, i.i.d. draws from any other continuous distribution on a compact interval and full support produces the same qualitative results:

either a threshold or interval audit rule (only in year 1) is optimal, and a roll-over rule in the interval $(\Delta_{\max \hat{x}}, 1]$ remains optimal. Finally, online appendix B.6 demonstrates a mechanism design approach to find the optimal audit rule.

5. CONCLUDING REMARKS

Agents routinely spend sizable portions of their annual budgets in the last month or even week of the fiscal year, before the budgets expire. There is mounting evidence that this use-it-or-lose-it spending is of low value to the principal. And this sort of waste appears to be widespread in public-sector organizations around the world. This article studies various instruments to limit such wasteful year-end spending. In short, rolling over unused funds to the next year should be allowed, and in some cases it might be optimal to "tax" the roll-over, so that only a fraction of saved funds are available next year. Audits and punishment to deter fund misuse are more costly options, and can be optimal for sufficiently large budgets (relative to likely spending needs), not too large audit costs, or for a sufficiently large cost of funds. Normally, absent roll-over, the optimal audit rule audits all spending above a specific spending threshold. If fund roll-over is allowed and effective, then the optimal audit rule changes, so that only a range of spending amounts in the interior of the budget set should be audited. Hence, allowing fund roll-over has implications for auditors and changes the shape of the optimal audit rule.

Some organizations allow their employees to spend some of their budget or expense account balance for private benefit, which is viewed as a perk and might make the organization more attractive as employer. This is consistent with the model in this article, by including these perks in the definition of legitimate spending needs and not classifying them as fund misuse. Accordingly, this article does not take a strong stance on what constitutes undesirable spending, but investigates how to minimize it once it is properly defined.

This article is only a first step in the analysis of optimal budget rules. Open questions remain. What are the optimal audit rules if there is weak punishment, so that auditing is not always an effective deterrent? How does the optimal roll-over rule change in an infinite-horizon setting where large savings might accumulate? How do the optimal budget rules change once a ratchet effect sets in? Beyond budgets, what is the optimal mechanism in this setting without institutional constraints? These are difficult questions, but the problem of wasteful spending is an important one and good answers can potentially save organizations a lot of money.

APPENDIX A: BACKGROUND CALCULCATIONS

A.1. *Principal Expected Utility (EU) Given Agent Reaction.* First, define the second-year EU as a function of the second year budget as

$$V(b_2) := \int_0^{\underline{a}_2} (\theta_2 + \lambda(b_2 - \underline{a}_2))\mathrm{d}G(\theta_2) + \int_{\underline{a}_2}^{b_2} (\theta_2 + \lambda(b_2 - \theta_2) - c_A)\mathrm{d}G(\theta_2) + \int_{b_2}^{u} (b_2 - c_A)\mathrm{d}G(\theta_2).$$

Now we can write out the principal EU given the agent reaction function over both years. If the roll-over rule is ineffective ($\hat{x} = 0$), the principal uses a threshold audit rule in both years:

$$\begin{aligned} EU = & \int_0^{\underline{a}_1} [\theta_1 + V(b + \Delta(b - \underline{a}_1)) + \lambda(1-\Delta)(b - \underline{a}_1)]\mathrm{d}G(\theta_1) \\ & + \int_{\underline{a}_1}^{b} [\theta_1 - c_A + V(b + \Delta(b - \theta_1)) + \lambda(1-\Delta)(b - \theta_1)]\mathrm{d}G(\theta_1) \\ & + \int_b^u [b - c_A + V(b)]\mathrm{d}G(\theta_1) - 2\lambda b. \end{aligned} \tag{A.1}$$

In the first line, the year 1 spending needs realize below the audit threshold, $\theta_1 \leq \underline{a}_1$, so the agent fulfills all spending needs and misuses funds until spending reaches the audit threshold, but no higher, to avoid punishment. The principal does not incur audit costs in year 1 as the audit threshold $\underline{a}_1$ is not exceeded. Since the agent does not spend everything, amount $\Delta(b - \underline{a}_1)$ is rolled over to year 2, and amount $(1 - \Delta)(b - \underline{a}_1)$ is returned to the principal at marginal utility λ.

In the second line, spending needs realize between the audit threshold and the available budget b. Hence, the agent fulfills the spending needs but does not misuse funds, since fulfilling spending needs already puts the spending above the audit threshold, and additional misuse would trigger punishment. The principal incurs audit costs. Since the agent does not spend everything, amount $\Delta(b - \theta_1)$ is rolled over to year 2.

In the third line, the spending needs exceed the budget b. Hence, the agent cannot fulfill all spending needs due to the budget constraint, but fulfills as many needs as possible by spending everything. The principal incurs audit costs (unless $\underline{a}_1 = b$). No funds are rolled over. The cost of funds over both years (without returns) is $2\lambda b$.

Next, if the roll-over rule is effective ($\hat{x} > 0$), then the principal uses an interval audit rule, as defined in Proposition 2, with $\underline{a}_1 \leq b - \hat{x} \leq \overline{a}_1$, in year 1, so the principal EU is

$$
\begin{aligned}
EU = &\int_0^{\underline{a}_1} [\theta_1 + V(b + \Delta(b - \underline{a}_1)) + \lambda(1 - \Delta)(b - \underline{a}_1)]\mathrm{d}G(\theta_1) \\
&+ \int_{\underline{a}_1}^{\overline{a}_1} [\theta_1 - c_A + V(b + \Delta(b - \theta_1)) + \lambda(1 - \Delta)(b - \theta_1)]\mathrm{d}G(\theta_1) \\
&+ \int_{\overline{a}_1}^{b} [\theta_1 + V(b + \Delta(b - \theta_1)) + \lambda(1 - \Delta)(b - \theta_1)]\mathrm{d}G(\theta_1) \\
&+ \int_b^u [b + V(b)]\mathrm{d}G(\theta_1) - 2\lambda b.
\end{aligned}
\tag{A.2}
$$

No audit cost is paid for large spending amounts exceeding $\overline{a}_1$, because the principal can be sure this spending is justified, since the agent wants to save these funds for roll-over unless a large θ_1-realization requires such high spending. Hence, misuse of funds does not have to be discouraged via audit, since the savings and roll-over incentives already achieve this, which saves the principal audit costs.

A.2. *Proofs.* Proof of Proposition 1. I will show that the agent reaction to any nonthreshold rule can be replicated with a threshold rule. Take any nonthreshold audit rule $R(s)$, and define r as $\min r \in \mathbb{R}$ such that $R(r) = 0$ and there exists no $s > r$ with $R(s) = 0$. Then, rule $R(s)$ leads to the same agent spending decisions as a threshold rule $A(s)$ with $\underline{a} = r$. This is because the agent best responds by spending $s_t = \underline{a} = r$ whenever $\theta_t \leq \underline{a} = r$ and $s_t = \theta_t$ whenever $\theta_t > \underline{a} = r$ under either rule.

If no minimum r as specified above exists, then use a threshold rule that also audits at the threshold, $A(s) = 1$ whenever $s = \underline{a}$, and is otherwise the same. Again take any nonthreshold audit rule $R(s)$, and define $r := \{\max r \in \mathbb{R} : R(r) = 1\}$. Then rule $R(s)$ leads to the same agent reaction as the threshold rule with $\underline{a} = r$. Hence, a threshold rule can always be part of the optimal policy. □

Proof of Proposition 2. Absent auditing, agents spend $s_1 = b - \hat{x}$ if $\theta_1 \leq b - \hat{x}$ by construction of $\hat{x}$, and $s_1 = \min\{\theta_1, b\}$ if $\theta_1 > b - \hat{x}$. To decrease fund misuse further for $\theta_1 \leq b - \hat{x}$, a continuous interval $s_1 \in (\underline{a}_1, \overline{a}_1)$ has to be audited, with $\underline{a}_1 < b - \hat{x}$ and $\overline{a}_1 > b - \hat{x}$ for incentive compatibility. This is the class of interval rules. An interval rule with $\underline{a}_1 = \overline{a}_1 = b - \hat{x}$ is a special case that does not audit. Hence, the optimal audit rule is in the family of interval rules.

In the next step, I narrow down what kind of interval rule is optimal. Setting $\underline{a}_1 < b - \hat{x}$ requires a specific $\overline{a}_1$ so that certain agent types do not misuse funds, that is, so that spending $s_1 = \underline{a}_1$ is incentive-compatible.

The difference in agent utility of spending $s_1 = \underline{a}_1$ versus spending $s_1 = \overline{a}_1$ if $\theta_1 \leq \underline{a}_1$, which needs to be nonnegative for those agents to choose $s_1 = \underline{a}_1$, is:

$$\int_{b+\Delta(b-\overline{a}_1)}^{b+\Delta(b-\underline{a}_1)} \theta_2 - (b + \Delta(b - \overline{a}_1))dG(\theta_2) + \int_{b+\Delta(b-\underline{a}_1)}^{u} \Delta(\overline{a}_1 - \underline{a}_1)dG(\theta_2) - \alpha(\overline{a}_1 - \underline{a}_1)$$
$$(A.3) \qquad = \frac{(\overline{a}_1 - \underline{a}_1)(\overline{a}_1\Delta^2 + \underline{a}_1\Delta^2 - 2b\Delta(1+\Delta) - 2\alpha u + 2\Delta u)}{2u} = 0,$$

a quadratic equation in $\overline{a}_1$, which is trivially fulfilled for $\overline{a}_1 = \underline{a}_1$. The other solution is

$$\overline{a}_1(\underline{a}_1) = \frac{2b\Delta(1+\Delta) + 2\alpha u - 2\Delta u - \underline{a}_1\Delta^2}{\Delta^2},$$

for which agents with $\theta_1 \leq \underline{a}_1$ are indifferent and hence $s_1 = \underline{a}_1$ is incentive-compatible. Moreover, it is easy to see that any $\theta_1 \geq b - \hat{x}$ spends $s_1 = \theta_1$ by construction of $\hat{x}$. And any $\theta_1 \in (\underline{a}_1, b - \hat{x})$ has a positive difference in expected utility between setting $s_1 = \theta_1$ and $s_1 = \overline{a}_1$ if (A.3) is fulfilled.

I will now show that interval rules for which (A.3) is negative are dominated by one for which it is zero. Such an alternative is constructed as follows: Determine $\underline{a}_1 < \overline{a}_1$ such that (A.3) holds. Keeping $\underline{a}_1$ fixed, reduce $\overline{a}_1$ slightly to $\overline{a}_1' < \overline{a}_1$ with $b - \hat{x} < \overline{a}_1'$. As a consequence, types $\theta_1 \leq \underline{a}_1$ switch from $s_1 = \underline{a}_1$ to $s_1 = \overline{a}_1'$. But there is a marginal type $\theta_1' \in (\underline{a}_1, b - \hat{x})$ who is indifferent between $s_1 = \theta_1'$ and $s_1 = \overline{a}_1'$. Any $\theta_1 > \theta_1'$ strictly prefers $s_1 = \theta_1$.

Under rule $(\underline{a}_1, \overline{a}_1')$ just constructed, any $\theta_1 < \theta_1'$ sets $s_1 = \overline{a}_1'$, since θ_1' is the marginal type, whereas any $\theta_1 \geq \theta_1'$ sets $s_1 = \theta_1$. The ex ante audit cost in year 1 is therefore $c_A(\overline{a}_1' - \theta_1')/u$, and the ex ante fund misuse in year 1 is

$$\int_0^{\theta_1'} \overline{a}_1' - \theta_1 dG(\theta_1) = \frac{\overline{a}_1'\theta_1' - \theta_1'^2/2}{u}.$$

Abusing notation, now consider instead the interval rule $\underline{a}_1 = \theta_1'$ and $\overline{a}_1 = \overline{a}_1'$. That is, the upper bound is the same, but the lower bound is larger, compared to the previous rule. Since θ_1' is the marginal type for whom $s_1 = \theta_1$ is weakly better than $s_1 = \overline{a}_1' = \overline{a}_1$, it follows that this rule sets (A.3) to zero. Consequently, all $\theta_1 \leq \underline{a}_1$ set $s_1 = \underline{a}_1$ as well, as their decision problem between $s_1 = \underline{a}_1$ and $s_1 = \overline{a}_1$ is the same as for $\theta_1 = \theta_1'$. And all $\theta_1 > \underline{a}_1$ set $s_1 = \theta_1$. Consequently, the ex ante audit cost under this rule is $c_A(\overline{a}_1 - \underline{a}_1)/u$, which is identical to the previous rule, because $\overline{a}_1 = \overline{a}_1'$ and $\underline{a}_1 = \theta_1'$. However, the ex ante fund misuse under this rule is smaller,

$$\int_0^{\underline{a}_1} \underline{a}_1 - \theta_1 \mathrm{d}G(\theta_1) = \frac{\underline{a}_1^2 - \underline{a}_1^2/2}{u} = \frac{\theta_1'^2 - \theta_1'^2/2}{u} < \frac{\overline{a}_1'\theta_1' - \theta_1'^2/2}{u},$$

since $\overline{a}_1' > \theta_1'$. Therefore, any interval rule for which (A.3) is negative is dominated by one for which it is zero. And clearly, any rule for which (A.3) is positive is suboptimal as well, as it could audit less and still achieve the same spending decisions by the agent. Hence, the interval rule that fulfills IC (A.3) with equality is best. □

PROOF OF PROPOSITION 3.

(i) First, I am going to show that $\hat{x}$ is maximized for Δ set to (3). Plugging the inverse of the uniform cumulative distribution function (CDF) into the expression for $\hat{x}$ in (1) yields

$$\Delta_{\max\hat{x}} = \arg\max_{\Delta\in[0,1]} \left(\frac{1}{\Delta} - \frac{\alpha}{\Delta^2}\right)u - \frac{b}{\Delta}.$$

Only consider $\Delta \geq \alpha$, since $\alpha > \Delta$ implies $\hat{x} < 0$ and hence $\hat{x} = 0$, and so can be ruled out as maximum. Clearly, for $\Delta \geq \alpha$, $\hat{x}$ is continuous, so by the maximum theorem a maximizing Δ exists on $[\alpha, 1]$. Differentiating with respect to Δ once yields

$$-\frac{u}{\Delta^2} + \frac{2\alpha u}{\Delta^3} + \frac{b}{\Delta^2}, \tag{A.4}$$

with an interior solution to the necessary first-order condition, and upper bound, at

$$\Delta_{\max\hat{x}} = \min\left\{\frac{2\alpha u}{u-b}, 1\right\},$$

which strictly exceeds α for all parameter values, and is positive due to the assumption of $b \leq u$. The objective $\hat{x}$ is not in general strictly concave in Δ, but it can still be shown that (3) is a maximum.

To show this, note both $\hat{x}$ as well as the first derivative is continuous for $\Delta > 0$. Moreover, setting derivative (A.4) to zero reduces it to a linear function, which implies there is only one solution to the first-order condition and this solution cannot be a saddle point. Furthermore, derivative (A.4) is strictly positive at $\Delta = \alpha$ for any $u \geq b$, so the solution to the first-order condition is a maximum. Consequently, (3) maximizes $\hat{x}$.

Second, if $\hat{x} > 0$, then an interval audit rule with $\underline{a}_1 \leq b - \hat{x}$ is used (Proposition 2). While keeping $\underline{a}_1$ fixed, a change in Δ can change $b - \hat{x}$, and consequently $\overline{a}_1(\underline{a}_1)$ fulfilling the incentive-compatibility constraint (A.3) changes. This Δ-change affects the expected audit cost and hence principal EU, whereas expected fund misuse remains the same when keeping $\underline{a}_1$ fixed.

Consider $\underline{a}_1$ such that $\overline{a}_1(\underline{a}_1) < b$. Taking the derivative of the principal EU in (A.2) with respect to Δ, using Leibniz' integral rule and simplifying, yields

$$\begin{aligned} &\int_0^{\underline{a}_1} \frac{\partial V(b+\Delta(b-\underline{a}_1))}{\partial b_2}\cdot\frac{\partial b_2}{\partial\Delta} - \lambda(b-\underline{a}_1)\mathrm{d}G(\theta_1) \\ &+\int_{\underline{a}_1}^{b} \frac{\partial V(b+\Delta(b-\theta_1))}{\partial b_2}\cdot\frac{\partial b_2}{\partial\Delta} - \lambda(b-\theta_1)\mathrm{d}G(\theta_1) - \frac{\partial\overline{a}_1(\underline{a}_1)}{\partial\Delta}c_A g(\overline{a}_1). \end{aligned} \tag{A.5}$$

Clearly, $\frac{\partial V(b_2)}{\partial b_2} = 1-(1-\lambda)G(b_2) \in [\lambda, 1]$, $\frac{\partial b_2}{\partial\Delta} = b - \underline{a}_1$ or $\frac{\partial b_2}{\partial\Delta} = b - \theta_1$, respectively. Moreover, based on (A.3), $\frac{\partial\overline{a}_1(\underline{a}_1)}{\partial\Delta} = \overline{a}_1' > 0$ if $\frac{\partial(b-\hat{x})}{\partial\Delta} > 0$, which in turn occurs iff $\Delta > \Delta_{\max\hat{x}}$, as the first part just showed. Similarly, $\overline{a}_1' < 0$ if $\Delta < \Delta_{\max\hat{x}}$, and $\overline{a}_1' = 0$ if $\Delta = \Delta_{\max\hat{x}}$. Plugging these in, the derivative becomes

$$\begin{aligned} &\int_0^{\underline{a}_1} (b-\underline{a}_1)[1-(1-\lambda)G(b_2)-\lambda]\mathrm{d}G(\theta_1) \\ &+\int_{\underline{a}_1}^{b} (b-\theta_1)[1-(1-\lambda)G(b_2)-\lambda]\mathrm{d}G(\theta_1) - \overline{a}_1' c_A g(\overline{a}_1), \end{aligned}$$

which is nonnegative for any $\Delta < \Delta_{\max\hat{x}}$, as $1-(1-\lambda)G(b_2) \geq \lambda$, so all three terms are weakly positive. At $\Delta = \Delta_{\max\hat{x}}$, the first two terms are weakly positive, whereas the last is zero, so the derivative is nonnegative. And for $\Delta > \Delta_{\max\hat{x}}$ sufficiently close to $\Delta_{\max\hat{x}}$, the last term is negative (since $\hat{x} > 0$ at $\Delta = \Delta_{\max\hat{x}}$ by assumption) whereas the first two are weakly positive. Consequently, this derivative is nonnegative for any $\Delta \leq \Delta_{\max\hat{x}}$, but can be zero and negative in $\Delta > \Delta_{\max\hat{x}}$, indicating a maximum in that range.

Third, consider $\underline{a}_1$ such that $\overline{a}_1(\underline{a}_1) > b$. In this case, the change of Δ does not change $\overline{a}_1(\underline{a}_1) > b$. Taking the derivative of (A.2) with respect to Δ, and simplifying as in the previous part yields

$$\begin{aligned}(A.6)\qquad &\int_0^{\underline{a}_1} \frac{\partial V(b+\Delta(b-\underline{a}_1))}{\partial b_2} \cdot \frac{\partial b_2}{\partial \Delta} - \lambda(b-\underline{a}_1)\mathrm{d}G(\theta_1) \\ &+ \int_{\underline{a}_1}^{b} \frac{\partial V(b+\Delta(b-\theta_1))}{\partial b_2} \cdot \frac{\partial b_2}{\partial \Delta} - \lambda(b-\theta_1)\mathrm{d}G(\theta_1) \geq 0.\end{aligned}$$

By assumption, $\overline{a}_1(\underline{a}_1) < b$ at $\Delta = \Delta_{\max\hat{x}}$. If, in addition, $\overline{a}_1(\underline{a}_1) > b$ at $\Delta = 1$, then an increase in $\overline{a}_1$ due to an increase in Δ can cause $\overline{a}_1 > b$, which in turn causes a discontinuous increase in expected audit costs of $c_A(u-b)/u > 0$ and a discontinuous drop in the principal EU of the same amount. This favors a $\Delta < 1$. Moreover, whether derivative (A.5) or (A.6) is valid depends on Δ. Taken together, principal EU is potentially decreasing in Δ when (A.5) is valid, weakly increasing when (A.6) is valid, and discontinuously decreasing when switching from (A.5) to (A.6). These calculations imply the optimal roll-over rule is $\Delta^* \in (\Delta_{\max\hat{x}}, 1]$.

Moreover, if $\hat{x} = b$ at $\Delta = 1$, then $\Delta^* = 1$ maximizes $\hat{x}$ and maximizes the roll-over, hence we can rule out any $\Delta < 1$ as optimal rule.

Note that $1-(1-\lambda)G(b_2)-\lambda$ becomes arbitrarily small as $\lambda \to 1$, whereas $\overline{a}_1' c_A g(b-\hat{x})$ is independent of λ (see (A.3) which determines $\overline{a}_1'$ and does not depend on λ). Consequently, the derivative (A.5) is negative at $\Delta >> \Delta_{\max\hat{x}}$, so a $\Delta^* \in (\Delta_{\max\hat{x}}, 1)$ is optimal.

(ii) Since $\hat{x}$ is maximized at $\Delta_{\max\hat{x}}$, $\underline{a}_1 \leq b - \hat{x}$ with $\hat{x} = 0$ at $\Delta = \Delta_{\max\hat{x}}$ implies $\hat{x} = 0$ and $\underline{a}_1 \leq b - \hat{x} = b$ for any Δ. In this case, taking the derivative of (A.1) with respect to Δ yields

$$\int_0^{\underline{a}_1} (b-\underline{a}_1)[1-(1-\lambda)G(b_2)-\lambda]\mathrm{d}G(\theta_1) + \int_{\underline{a}_1}^{b} (b-\theta_1)[1-(1-\lambda)G(b_2)-\lambda]\mathrm{d}G(\theta_1) \geq 0$$

due to $1-(1-\lambda)G(b_2) \geq \lambda$. Hence, $\Delta^* = 1$ is optimal in this case, though may not be the only optimum. If $\Delta_{\max\hat{x}} = 1$, then as shown above, principal EU cannot decrease in Δ for $\Delta < \Delta_{\max\hat{x}}$, so $\Delta^* = 1$ is optimal. If $\hat{x} = b$ at $\Delta = 1$, then $\Delta = 1$ maximizes $\hat{x}$ even if $\Delta_{\max\hat{x}} < 1$ since the min-operator in $\hat{x}$ is binding. Consequently, $\Delta^* = 1$ is optimal. Finally, if $\overline{a}_1(\underline{a}_1) > b$ at $\Delta < \Delta_{\max\hat{x}} < 1$, then $\overline{a}_1' = 0$ and hence the derivative in (A.5) is nonnegative, hence $\Delta^* = 1$ is optimal.

□

PROOF OF PROPOSITION 4.

(i) Using Leibniz' integral rule, the marginal principal EU of changing the threshold $\underline{a}_2$ is

$$\begin{aligned}\frac{\partial EU}{\partial \underline{a}_2} = \frac{\partial V(b_2)}{\partial \underline{a}_2} &= (\underline{a}_2 + \lambda(b_2 - \underline{a}_2))g(\underline{a}_2) + \int_0^{\underline{a}_2} -\lambda \mathrm{d}G(\theta_2) - (\underline{a}_2 + \lambda(b_2-\underline{a}_2) - c_A)g(\underline{a}_2) \\ (A.7)\qquad &= -\lambda G(\underline{a}_2) + c_A g(\underline{a}_2).\end{aligned}$$

The maximization problem for $\underline{a}_2$ is strictly concave, as using the uniform probability density function (PDF) and CDF and differentiating the marginal utility in (A.7) yields

$-\lambda/u < 0$. Hence, a unique interior solution is guaranteed, and using the uniform PDF and CDF in (A.5) and rearranging yields

$$\underline{a}_2^* = \frac{c_A}{\lambda}.$$

If $b = u$ or $b < b_2$, then there is no discontinuity at $\underline{a}_2 = b$, since there is no probability mass point at $s_2 = b$ given the agent reaction function. Hence, by strict concavity, the interior solution is the optimal policy, unless the constraint $c_A/\lambda \le b$ is binding, in which case the corner solution is optimal.

However, if $b_2 = b < u$, then there is a discontinuity at $\underline{a}_2 = b$, since all $\theta_2 \in [b, u]$ agent types spend $s_2 = b$ due to the budget constraint. So even if $c_A/\lambda \le b$ is not binding, the corner solution can be optimal due to the discontinuity. The interior solution (given $c_A/\lambda \le b$) is optimal in this case if and only if

$$\int_0^{\underline{a}_2^*} \theta_2 + \lambda(b_2 - \underline{a}_2^*)\mathrm{d}G(\theta_2) + \int_{\underline{a}_2^*}^{b_2} \theta_2 + \lambda(b_2 - \theta_2) - c_A\mathrm{d}G(\theta_2) + \int_{b_2}^{u} b_2 - c_A\mathrm{d}G(\theta_2)$$

$$\text{(A.8)} \qquad \ge \int_0^{b_2} \theta_2\mathrm{d}G(\theta_2) + \int_{b_2}^{u} b_2\mathrm{d}G(\theta_2) \iff \frac{b_2^2\lambda}{2} + \frac{c_A^2}{2\lambda} \ge uc_A,$$

where the left-hand side is the expected principal utility in year 2 from using the interior solution $\underline{a}_2^* = c_A/\lambda$ and the right-hand side is the principal expected utility from using $\underline{a}_2 = b$.

(ii) If, instead, (A.8) is not fulfilled, then the corner solution $\underline{a}_2 = b$ is optimal.

□

Proof of Proposition 5. Since the year 2 audit threshold is independent of both $\underline{a}_1$ and Δ, as it can condition on b_2, the optimal $\underline{a}_2^*$ in Proposition 4 is optimal for all $\underline{a}_1$ and Δ.

(i) If $\alpha > 1/2$, then $\Delta_{\max\hat{x}} = 1$, hence $\Delta^* = 1$ according to Proposition 3. Similarly, $\Delta_{\max\hat{x}} \ge 1 \iff b \ge u(1 - 2\alpha)$. Since $b \ge u(1 - 2\alpha)$ is fulfilled whenever $\alpha > 1/2$, the former condition is not separately needed.

(ii) Fix u and α. If b is sufficiently large (relative to u) and c_A/λ is sufficiently small (relative to b), then $\underline{a}_1^t = \underline{a}_2^* = c_A/\lambda$ is optimal for any $\Delta > \alpha$ according to Propositions A.6 and 4. Moreover, for b large enough, $\hat{x} = 0$ for any $\Delta \in [0, 1]$, so the year 1 audit rule is a threshold rule, and hence $\Delta^* = 1$ is optimal according to Proposition 3(ii). Since $\hat{x} = 0$ at $\Delta = 1$ whenever $b > (1 - \alpha)u$, this condition requires b to be large relative to u or $-\alpha$.

(iii) Fix u. Small α and b imply $\hat{x} > 0 \iff (1 - \frac{\alpha}{\Delta})u > b$, so the year 1 audit rule is an interval rule. If in addition $b > \frac{(1-\alpha)u}{2}$, then $\hat{x} < b$ at $\Delta = 1$. Small α also implies $\Delta_{\max\hat{x}} < 1$, hence $\Delta^* \in (\Delta_{\max\hat{x}}, 1]$ is optimal if $\overline{a}_1(\underline{a}_1^i) < b$ according to Proposition 3(i). Moreover, a sufficiently small c_A/λ and hence $2c_A/\lambda$ (relative to b) implies active auditing is optimal in both years (Propositions 4 and A.7), that is, $\underline{a}_1^i \le 2c_A/\lambda$, $\underline{a}_2^* = c_A/\lambda < b$.

(iv) Fix u. If $b \le \frac{(1-\alpha)u}{2}$, then $\hat{x} = b$ at $\Delta = 1$. Hence, no $\Delta < 1$ can increase $\hat{x}$, thus $\Delta = 1$ is optimal (Proposition 3). Since $\hat{x} > 0$, an interval audit rule is optimal in year 1. Moreover, since $\hat{x} = b$, the agent does not misuse funds for any realization θ_1, thus no auditing in year 1 is optimal. In year 2, either active auditing ($\underline{a}_2^* = c_A/\lambda$) if c_A/λ is sufficiently small (relative to b) or no auditing ($\underline{a}_2^* = b$) if c_A/λ is sufficiently large (relative to b) is optimal (Propositions A.7 and 4).

(v) Fix u. As in (iii), small α and b imply $\hat{x} > 0$, so the year 1 audit rule is an interval rule. A large c_A/λ relative to b implies $\underline{a}_1^i = b - \hat{x}$, $\underline{a}_2^* = b$, which also implies $\overline{a}_1(\underline{a}_1) =$

$b - \hat{x} < b$. If $b > \frac{(1-\alpha)u}{2}$, then $\hat{x} < b$ at $\Delta = 1$. Small α also implies $\Delta_{\max\hat{x}} < 1$, hence $\Delta^* \in (\Delta_{\max\hat{x}}, 1]$ is optimal according to Proposition 3(i).

□

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information

REFERENCES

Baumann, S.,"Putting It Off for Later: Procrastination and End of Fiscal Year Spending Spikes, " *Scandinavian Journal of Economics* 121 (2019), 706–35.

Ben-Porath, E., E. Dekel, and B. L. Lipman, "Optimal Allocation with Costly Verification, " *American Economic Review* 104 (2014), 3779–813.

Bird, D., and A. Frug, "Dynamic Non-Monetary Incentives, " *American Economic Journal: Microeconomics* 11 (2019), 111–50.

Border, K. C., and J. Sobel, "Samurai Accountant: A Theory of Auditing and Plunder, " *Review of Economic Studies* 54 (1987), 525–40.

Brimberg, J., and W. J. Hurley, "Allocating Operating Funding in the Public Sector and the Newsvendor Problem, " *Journal of the Operational Research Society* 66 (2015), 1035–43.

Digiday, "'Irrational Budget Dumping': End of the Year Means Use-It-or-Lose-It Marketing Projects," 2017. https://digiday.com/marketing/irrational-budget-dumping-end-year-means-use-lose-marketing-projects/, accessed 10/8/2021.

Freixas, X., R. Guesnerie, and J. Tirole, "Planning under Incomplete Information and the Ratchet Effect, " *Review of Economic Studies* 52 (1985), 173–91.

Harris, M., and A. Raviv, "The Capital Budgeting Process: Incentives and Information, " *Journal of Finance* 51 (1996), 1139–74.

———, "Capital Budgeting and Delegation, " *Journal of Financial Economics* 50 (1998), 259–89.

HM Treasury, "Consolidated Budgeting Guidance: 2021-22," 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/969175/2021-22_CBG_For_Publication.pdf.

Hurley, W., J. Brimberg, and B. Fisher, "Use It or Lose It: On the Incentives to Spend Annual Defence Operating Budgets, " *Defence and Peace Economics* 25 (2014), 401–13.

Jones, L. R.,"Outyear Budgetary Consequences of Agency Cost Savings: International Public Management Network Symposium, " *International Public Management Review* 6 (2005), 139–68.

Khalil, F., D. Kim, and J. Lawarree, "Use It or Lose It, " *Journal of Public Economic Theory* 21 (2019), 991–1016.

Li, Y.,"Mechanism Design with Costly Verification and Limited Punishments, " *Journal of Economic Theory* 186 (2020), 105000.

Liebman, J. B., and N. Mahoney, "Do Expiring Budgets Lead to Wasteful Year-End Spending? Evidence from Federal Procurement, " *American Economic Review* 107 (2017), 3510–49.

Malenko, A.,"Optimal Dynamic Capital Budgeting, " *Review of Economic Studies* 86 (2019), 1747–78.

McPherson, M. F., "An Analysis of Year-End Spending and the Feasibility of a Carryover Incentive for Federal Agencies," Technical Report, Naval Postgraduate School, Monterey, CA, 2007.

Military Times, "Use-It or Lose-It: DoD Dropped $4.6 Million on Crab and Lobster, and $9,000 on a Chair in Last-Minute Spending Spree," 2019. https://www.militarytimes.com/news/your-military/2019/03/12/use-it-or-lose-it-dod-dropped-46-million-on-crab-and-lobster-and-9000-on-a-chair-in-last-minute-spending-spree/, accessed 10/8/2021.

Miller, G. J., D. Robbins, and J. Keum, "Incentives, Certification, and Targets in Performance Budgeting, " *Public Performance & Management Review* 30 (2007), 469–95.

Roper, A. H., and M. E. Ruckes, "Intertemporal Capital Budgeting, " *Journal of Banking & Finance* 36 (2012), 2543–51.

Townsend, R. M.,"Optimal Contracts and Competitive Markets with Costly State Verification, " *Journal of Economic Theory* 21 (1979), 265–93.