Dynamic Compensation Contracts with Private Savings

Zhiguo He
Booth School of Business, University of Chicago

This article studies a dynamic agency problem in which a risk-averse agent can save privately. In the optimal contract, (i) cash compensations exhibit downward rigidity to failures; (ii) permanent pay raises occur when the agent's historical performance is sufficiently good; (iii) and when the agent is dismissed due to poor performance, he walks away with severance pay to support his post-firing consumption at the current compensation level. Thus, the theory can simultaneously explain the popularity of options-like compensation contracts and the increasing incidence of forced turnovers with sizeable severance pay. (*JEL* D86, J31, J33)

1. Introduction

In the past three decades, stock options as a form of executive compensation have become extremely popular (e.g., Murphy 1999). The options-type remuneration contract offers downside protection in cash compensation, which makes managers less averse to negative performances. In the meantime, in recent years forced turnovers have increased among executives (e.g., Kaplan and Minton 2008), but often with sizable severance pays (e.g., Yermack 2006). The literature often treats these facts as separate phenomena, which is somewhat unsatisfactory.

We propose a dynamic agency model with private saving that generates all three stylized facts mentioned above. In the model, the risk-averse agent controls the firm's profitability through unobservable actions, and he can save privately (or secretly save, have hidden savings). Because the agent can save privately to undo any compensation contract that punishes him severely following poor performance, the optimal contract offers downside-protected compensation packages (i.e., downward rigid) in order to mitigate the agent's undoing activities. However, this gives too many "carrots" to the agent. To maintain proper working incentives, the compensation policy should invoke "sticks" more often, which would result in more frequent forced turnovers

I would like to thank my advisers Mike Fishman and Arvind Krishnamurthy at Northwestern University for their endless support. Discussions with Peter DeMarzo, Doug Diamond, Jiro Kondo, Debbie Lucas, Adriano Rampini, Yuliy Sannikov, Paola Sapienza, and Andrzej Skrzypacz are gratefully acknowledged. All errors are mine. Send correspondence to Zhiguo He, University of Chicago, Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637; telephone: (773) 834-3769. E-mail: zhiguo.he@chicagobooth.edu.

doi:10.1093/rfs/hhs054 Advance Access publication March 19, 2012

following poor performance. And, severance pay is necessary to avoid the agent's ex ante undoing activities during his tenure on the job.

In the dynamic contracting literature, Rogerson (1985) among others has found that with contractible savings—that is, when the principal can dictate the agent's consumption/saving decisions—the optimal consumption pattern tends to be *front loaded*. That is, the so-called inverse-martingale property implies that the agent's expected marginal utility from consumption increases over time. With access to a private savings account, the agent will smooth his consumption, thereby devastating the incentive scheme designed in a front-loaded contract. Consequently, the optimal contract derived in this article features a back-loaded consumption pattern, and the agent's motive to save privately is absent.

The optimal contracting problem with private savings is complicated in general. In the specific setting considered in this article, cash flows follow a Poisson process. The cash-flow arrival intensity is controlled by the agent's three levels of unobservable effort (action): shirking, working, and myopic, and the optimal contract implements the interior working effort. Shirking leads to no cash flow in the next time interval, while working generates a positive success intensity. A myopic action helps improve the short-term "hard" cash-flow performance, but it hurts the firm's long-run value. We envision that the long-run destruction, usually taking forms unforeseen by investors, will be realized after the agent's tenure and is therefore not contractible.[1] Therefore, discouraging myopic behavior requires the optimal contract to avoid incentives that load excessively on short-term cash-flow performance.

This requirement, together with the linear effort cost structure, implies that the optimal contract provides the exact working incentives (i.e., a binding incentive-compatibility constraint) for the agent against shirking. As a result, the agent loses nothing by shirking. In addition, because the agent can also save privately, the contract cannot specify a consumption cut after the agent's poor performance. The argument is based on the agent's potential joint deviation of "shirking and saving." Imagine a counterfactual situation where a contract assigns decreasing consumption following no success. Because the agent loses nothing by shirking, and under shirking the path of no success occurs with probability one, the shirking agent who saves concurrently can strictly improve his payoff by smoothing his consumption along the path of no success. As a result, investors cannot punish the agent by cutting his consumption after poor performance. Instead, in our model, the optimal contract resorts to termination as the "stick" for incentive provision.

[1] This captures the cost of high-powered incentive schemes, a well-documented economic phenomenon (e.g., Levitt and Dubner 2005; Larkin 2006). In the literature of corporate finance, this idea is connected to Stein (1989) and the ongoing literature on overvalued equity and related agency issues (e.g., Jensen 2005; Efendi, Srivastava, and Swanson 2007). One of the most celebrated examples, cited from Larkin (2006), is Sears' experience of offering commissions to its auto mechanics based on total charges for parts and labor. Mechanics responded to this scheme by ordering unneeded repairs, and Sears ended up settling a class-action lawsuit over excessive billing.

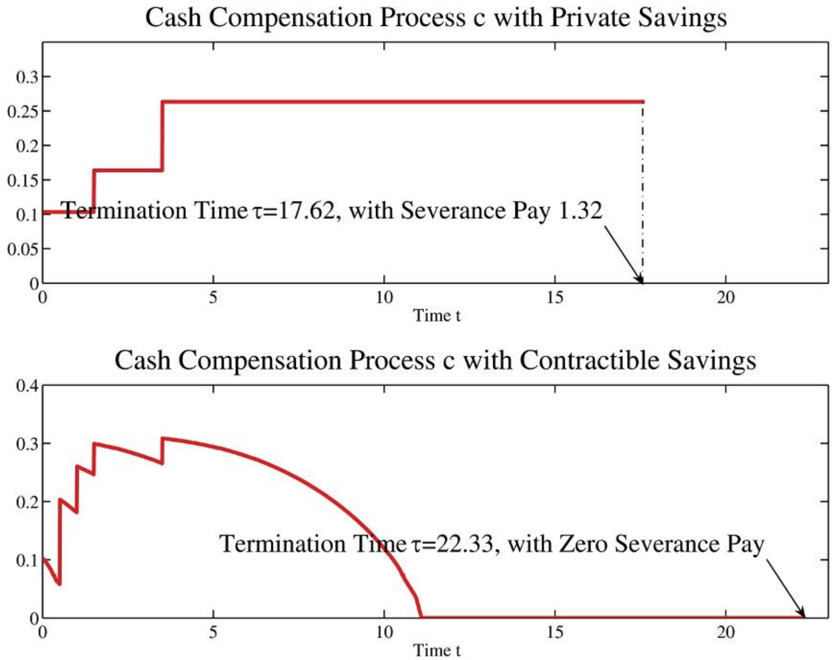

Figure 1
Optimal consumption patterns with private savings (top panel) and contractible savings (bottom panel)

We solve for the optimal contract based on the recursive method. In a natural implementation of the optimal contract, the agent's wages (which correspond to his cash compensation as well as consumption) are downward rigid; the agent is guaranteed with the current pay level, and works for future pay raises (promotions). The agent is dismissed after a streak of poor performance, and therefore loses the chance of future promotions. Nevertheless, the agent walks away with a severance pay that supports his post-firing consumption at the current compensation level.

Whether private savings are possible or not makes the optimal contract drastically different. Figure 1 compares our model with another one of the same setting except that the agent's savings are contractible. For both models, the agent starts with the same initial state, and experiences the same cash-flow performance (at $t = 0.5$, 1, 1.5, and 3.5). When savings are contractible, the agent's consumption displays a "zig-zag" pattern, responding actively to not only cash-flow realizations (successes in this model) but also no-cash-flow realizations (failures in this model). In contrast, when the agent can save privately, his consumption under the optimal contract is adjusted upward only, and never responds to poor performance. Regarding severance pay, the fired agent leaves the company with nothing in the case of contractible savings.

However, the private-saving case features a positive severance pay when the agent is dismissed due to his poor performance, and forced turnovers occur more often than in the contractible savings case.[2] Intuitively, in the private savings case, a harsher termination policy counterbalances the more lenient cash compensation policy in order to maintain proper working incentives for the agent.

It is important to stress that, as we will mention in Remark 3 in Section 3.2, the strict downward rigidity hinges on the assumption of zero success probability under shirking.[3] In general, compared with models with contractible savings, the private-saving technology makes the optimal compensation pattern more rigid in an agent's poor performance, which resembles the asymmetry in options payoff. And, although cash compensation is rigid in our model, the agent's continuation payoffs provide working incentives through future performance-based promotions or firing. Finally, the severance pay is increasing in the agent's past performance, and decreasing in his outside option.

We discuss various empirical predictions based on our theory in Section 6.2. We suggest that empirical researchers pay attention to the wedge between the (cash) incentives due to positive shocks and those due to negative shocks, a measure that is presumably increasing with the usage of stock options in compensation packages. The central prediction of our theory is that managers who can easily smooth out their on-the-job compensation incentives will receive cash compensations that are less sensitive to their downward performance, and our model suggests that low-corporate-governance firms should have a greater cash-incentive wedge on their compensation policies. Finally, as suggested by Figure 1, our theory predicts a positive relationship between the use of options-like contracts and forced turnovers. This prediction can be readily tested based on available data, and the answer may differentiate our theory from the standard entrenchment story (e.g., Bebchuk and Fried 2004).

Literature review This article belongs to the burgeoning continuous-time contracting literature.[4] We use a framework similar to that of DeMarzo and Sannikov (2006), who study a continuous-time version of the DeMarzo and

[2] There are other articles showing that positive severance payments, by soothing the agent's fear of dismissal, might provide proper incentives for risk-taking (Berkovitch, Israel, and Spiegel 2000), or complete information disclosure (Eisfeldt and Rampini 2008). In essence, these findings are along the same line as this article: By promising a generous severance package, the contract prevents the agent from harmful deviation strategies (e.g., shirking and saving in this article; see Section 3.2).

[3] This point can be seen in the fourth paragraph in the Introduction when we explain the joint deviation of "shirking and saving." Suppose that the success probability under shirking remains strictly positive, and in the optimal contract the consumption following a success rises. If the consumption drops sufficiently slowly along the no-success path, then under shirking the marginal utility might be non-increasing in expectation (in contrast, if shirking leads to no success for sure, then the agent's marginal utility is always increasing). As a result, "shirking and saving" might not improve the agent's value. See Remark 3 in Section 3.2 for more details.

[4] This literature builds on the vast literature on discrete-time long-term agency models (Spear and Srivastava 1987; Phelan and Townsend 1991; etc.).

Fishman (2007) model. Biais et al. (2007) show that the contract in DeMarzo and Sannikov (2006) arises in the limit of the discrete-time model. In all three articles the agent is risk-neutral, which eliminates the saving incentives.[5] This article is also closely related to Sannikov (2008), who studies an optimal contracting problem with a risk-averse agent in which savings are observable and contractible.

The problem studied in this article is akin to the literature on unemployment insurance—for example, Hopenhayn and Nicolini (1997). Kocherlakota (2004) solves an optimal unemployment insurance contract under the setting of private savings with a single success of permanent employment. In the current article, the key assumption of implementing interior effort under a linear effort cost structure is the same as Kocherlakota (2004). However, Mitchell and Zhang (2007) show that it is never optimal to implement interior effort in the setting of Kocherlakota (2004). Via identifying the agent's most profitable deviation strategy, Mitchell and Zhang (2007) provide a nice solution to optimal contracting with private savings and binary effort choices.

Our article differs from Kocherlakota (2004) along several dimensions. First, the interior effort is indeed optimal in our setting. This is because the excessive "myopic" action leads to long-run detrimental effects to the firm, and this new element allows us to provide rigorous justifications and proofs for the optimality of the contract. Second, because in our model the agent enjoys positive perks only during his tenure, the optimal contract features an endogenous termination with severance pay in the employment contract. This feature is absent in Kocherlakota (2004). Finally, we use the technique in Sannikov (2008) to study a more general setup with multiple cash flows in a continuous-time framework.[6,7]

The rest of this article is organized as follows. Section 2 describes the model. Section 3 and Section 4 solve the relaxed problem recursively. In Section 5, we verify that the solution to the relaxed problem is indeed the solution to

---

[5] The follow-up studies include He (2009), who studies executive compensation by analyzing a geometric Brownian motion model, and Piskorski and Tchistyi (2010), who study optimal mortgage design by considering exogenous regime switching in the investors' discount rate. Another strand of continuous-time contracting literature starts from Holmstrom and Milgrom (1987). This framework allows for private savings, due to the absence of wealth effect. See, for example, Fudenberg, Holmstrom, and Milgrom (1990), Williams (2006), and He (2011); the latter two characterize the optimal contract with private savings.

[6] Other related literature on agency issues with access to credit market (especially hidden savings) includes Allen (1987), Bizer and DeMarzo (1999), Cole and Kocherlakota (2001), Werning (2001), and Bisin and Rampini (2006). Fundamentally, the issue of hidden savings is that *hidden information* (in contrast to *hidden action* as effort) arises during the long-term contractual relationship. Under a discrete state-space framework, Fernandes and Phelan (2000) and Doepke and Townsend (2006) propose a recursive method to handle this issue for a certain class of problems.

[7] Harris and Holmstrom (1982) find that the downward-rigid wage is optimal. Their mechanism is fundamentally different from ours. In their learning-based model, without moral hazard issues, the first-best wage contract features a constant wage for the risk-averse agent to fully insure his productivity shocks. If the agent can quit, then a competitive labor market implies that looking forward, the agent's future compensation has to stay above his expected productivity at any time during the employment. In other words, the agent's ex post participation constraint might be binding. As a result, to match the agent's outside option, the contract will specify a wage raise in response to sufficiently good news about the agent's productivity.

the original problem. Section 6 discusses the optimal contract and considers several extensions. We conclude in Section 7. All proofs are given in the Appendix.

2. The model

2.1 Technology

Consider a continuous-time infinite-horizon principal-agent model, where the risk-neutral investors (the principal) hire a risk-averse agent for business operation. For any $t > 0$, the firm generates cash flows YdN_t during the interval $(t - dt, t]$, where $\{N_t\}$ is a standard Poisson process with intensity $\{a\}$, and Y is a positive constant. Later we use "cash flow," "jump," and "success" interchangeably. The cash flows are observable and contractible.

Denote by $\mathcal{F}^N = \left\{\mathcal{F}_t^N\right\}_{t\geq 0}$ the filtration generated by $\{N_t\}$. Our analysis is based on the stochastic calculus in jump processes (e.g., Protter 1990), in which the following notation is required. For any $\mathcal{F}^N$-adapted right-continuous-left-limit (RCLL) process $\{A\}$, define its left-hand limit as $A_{t-} \equiv \lim_{s\uparrow t} A_s$, which is $\mathcal{F}^N$-predictable. Essentially, A_{t-} (A_t) is the process $\{A\}$'s time-t value before (after) observing whether or not there is a cash-flow realization during the interval $(t - dt, t]$.

The agent can generate at most K cash flows. Though our results hold for any finite K (because we use induction analysis), for the sake of convenience we present results for the stationary case where $K \to \infty$. When employment ceases—that is, the agent is fired—investors can liquidate the firm's assets for an exogenous value L, if the agent does not take myopic action (to be discussed shortly). The liquidation value L is below the first-best asset value, which implies that early termination is inefficient. One can easily endogenize L through a costly replacement with another new agent. Both the agent and investors discount future payoffs at a constant market interest rate $r > 0$.

The agent's unobservable effort controls the intensity of the jump process. Specifically, the agent's effort process $\{a\}$ is $\mathcal{F}^N$-predictable—that is, making effort choice before knowing whether or not a cash flow occurs at that instant. There are three effort levels—that is, $a_t \in \{0, p, \overline{p}\}$, where $\overline{p} > p > 0$ and $\overline{p} - p = \epsilon$ is small. The agent's nonpecuniary personal effort cost when exerting a_t, in terms of the agent's utilities, is $b\left(\frac{a_t}{p} - 1\right)dt$, where b is a positive constant.[8] We call the lowest effort $a_t = 0$ *shirking*. By shirking, the agent enjoys a private benefit bdt (a negative personal cost), but the intensity of cash flow is zero.[9] The agent can also choose the *working* effort $a_t = p$. In

8 The discrete structure of the agent's action space is immaterial. The key is the linear structure of the agent's effort cost, and implementing the interior effort in the optimal contract. For instance, the analysis will be the same if we assume that $a_t \in [0, \overline{p}]$, and there exists a critical level $p < \overline{p}$ such that any $a_t \in (p, \overline{p}]$ triggers the myopic loss.

9 The assumption that shirking leads to zero cash-flow intensity is crucial for the perfect downward rigidity in compensation patterns derived in the optimal contract. See the discussion in Remark 3 in Section 3.2.

this case he obtains zero private benefit, but the firm generates cash flows with a probability pdt in the interval $(t-dt, t]$.

In this model, the agent may exert the highest *myopic* effort $\overline{p} > p$ to increase the cash flow intensity. In the spirit of Stein (1989), this myopic action is detrimental to the liquidation value L, as it represents the short-term performance-enhancing strategies that hurt the firm's long-run value. We take a reduced-form approach by assuming that the present value of these losses borne by investors is Δdt, where Δ is a positive constant.

More importantly, we assume that these long-run losses are noncontractible.[10] There are other ways to interpret these noncontractible losses, and in this article we keep our interpretations general. Note that L can be interpreted as the firm's ongoing value after the agent is fired, and the loss due to these myopic actions can be uncovered only after the agent's tenure.[11] This idea is also similar to the multitasking problem studied in Holmstrom and Milgrom (1991) (see the related analysis in Section 7.2). There, if the compensation contract imposes excessive incentives on easy-to-measure hard performance (cash-flow occurrence in this model), the agent will ignore other dimensions of soft performances that are critical to the firm—for instance, refusing to collaborate with his colleagues and thereby lowering their efficiency. The bottom line is that the myopic action captures the cost of high-powered incentive schemes, a well-documented fact in both economic and finance literatures (e.g., Stein 1989; Jensen 2005; Larkin 2006; etc.).

Throughout the article we consider the case where it is optimal to implement the working effort $a_t = p$ always. We verify the optimality of this policy in Section 5. Because in equilibrium the myopic action is never invoked, the liquidation value L is a constant, and without loss of generality we set $L = 0$.

2.2 The agent

Utility function. The agent's instantaneous utility from consumption is $u(\widehat{c}_t)$, where $u > 0$, $u' > 0$, $u'' \leq 0$, and $\widehat{c}_t \geq 0$ is the consumption rate. When the agent is hired by the firm, his total utility $\widetilde{u}(\widehat{c}_t, a_t)$ takes an additive form:

$$\widetilde{u}(\widehat{c}_t, a_t) = u(\widehat{c}_t) + b\left(1 - \frac{a_t}{p}\right). \tag{1}$$

[10] As argued in note 4 in Hart and Moore (1998), if investors value these liquidated assets more than the market does, then the liquidation value can be nonverifiable, therefore noncontractible. We can also formally model this idea in the following way. Assume that the liquidation value L is positive and random, and whenever the agent exerts $a = \overline{p}$, the expected (discounted) liquidation value L drops by at least Δdt. During liquidation, investors (as banks with specialty in locating the second-best users) handle the liquidation process, and report a liquidation value $\widehat{L}$ that might differ from the true liquidation value L. Ruling out a third party (due to the possibility of collusion, etc.), the information revealed by the report $\widehat{L}$ becomes as if noncontractible.

[11] For instance, in August 2007, Dell restated down its past four years' earnings by up to $150 million, and the executives who were responsible to this scandal had left the company. (Source: "Dell to Restate Earnings After Probe," http:biz.yahoo.com/ap/070816/dell_restatement.html.)

When the agent is unemployed, his instantaneous utility is simply $u(\widehat{c}_t)$ without the effort-dependent term.

Remark 1. The agent's post-firing utility $u(\widehat{c}_t)$ is below his total utility $u(\widehat{c}_t) + b$ if he is shirking inside the firm. The underlying assumption of this specification is that the shirking benefit—which can be interpreted as enjoying perks or even personal satisfaction—is available only when the agent is employed by the firm. This constitutes one essential difference from Kocherlakota (2004): in that model, the agent has the same utility function independent of whether he is in or out of the unemployment insurance program. Therefore, in our model, termination is a punishment mechanism and is invoked along the equilibrium path, while termination never occurs in Kocherlakota (2004).

For the working effort to be optimal, we have to rule out "extreme" wealth effects. Formally, we assume that there exists a strictly positive number γ_L such that

$$\inf_{\widehat{c}_t \geq 0} u'(\widehat{c}_t) = \gamma_L > 0. \tag{2}$$

Intuitively, from the agent's view, the monetary equivalence (marginally) of the effort cost is b/u'. Therefore, condition (2) places an *upper* bound on the agent's monetary cost of effort.[12]

Though our results hold for general utility functions (see Section 7.3), in the main analysis we focus on a special form of $u(\cdot)$, which is the modified CARA (constant absolute risk aversion) utility defined as follows:

$$u(\widehat{c}_t) = \begin{cases} 1 - e^{-\gamma \widehat{c}_t} & \text{if } \widehat{c}_t < \frac{1}{\gamma} \ln \frac{\gamma}{\gamma_L} \\ 1 - \frac{\gamma_L}{\gamma} + \gamma_L \left(\widehat{c}_t - \frac{1}{\gamma} \ln \frac{\gamma}{\gamma_L} \right) & \text{otherwise} \end{cases}. \tag{3}$$

In words, to respect condition (2), we replace the upper part (when $\widehat{c}_t \geq \frac{1}{\gamma} \ln \frac{\gamma}{\gamma_L}$) of the CARA utility by a linear function with slope $\gamma_L > 0$ (so the agent becomes risk-neutral with a marginal utility of γ_L when his consumption is sufficiently high). The CARA form possesses a convenient feature that the marginal utility is linear in the utility level, which simplifies our analysis. See Section 7.3 for the analysis of general utility functions. Note that the additive form in Equation (1) implies wealth effect, and it is different from Holmstrom and Milgrom (1987), who assume CARA utility but monetary effort costs.[13]

---

12 Given a finite number (K) of cash-flow jumps, the marginal utility level γ_L may never be reached in equilibrium.

13 The monetary effort cost specification means that, given consumption $\widehat{c}_t$ and action a_t, the agent's instantaneous utility is $1 - \exp\left(-\gamma\left(\widehat{c}_t + b\left(1 - \frac{a_t}{p}\right)\right)\right)$. This differs from specification (1) in a substantial way.

Private savings. In this article the agent can save privately for consumption-smoothing purposes. As first noted by Rogerson (1985), if the agent's utility is additive, as in Equation (1), the optimal contract without private savings features an "inverse-martingale property." Under this property, the agent's marginal utility from consumption follows a submartingale (i.e., the expected marginal utility increases over time). This is against the "consumption-smoothing property" if the agent can save privately.

We rule out the agent's borrowing from a third party. Borrowing opportunities where a bank expects the agent to pay back his loan are inconsistent with the agents' private savings, unless the bank has certain technologies to enforce the repayment (see note 14). Our main results go through if the borrowing rate exhibits a sufficiently large spread over the saving rate r, or if the agent faces a fixed borrowing limit.[14]

2.3 Employment contract

An employment contract specifies a cash compensation (not consumption) process $\{c_t \geq 0 : 0 \leq t < \tau\}$ and a lump-sum transfer $F_\tau \geq 0$, where τ is the endogenous termination time when the agent is fired. We denote such a contract by $\Pi \equiv \{\{c\}, F_\tau, \tau\}$, and each element is $\mathcal{F}^N$-adapted (i.e., performance-based compensation contract). Here, because of the agent's limited liability, any contractual payment to the agent must be nonnegative.

In this abstract employment contract, we can interpret the cash compensation c_t as wages that the agent receives during his employment, and F_τ as the severance pay when the agent is dismissed. The combination of wages and severance pay can be understood as an implementation of the cash compensation contract. Keep in mind that in reality there may be other compensation packages (say performance-based vesting in stock and option grants) to implement the same cash compensation policy (we come back to this point in Section 6.1.3).

The agent has zero initial wealth. For simplicity, we assume that after τ the agent remains unemployed forever (so his outside option is zero). However, we will see that due to the possibility of private savings, the agent will maintain an endogenous consumption level after he is fired.

14 With a fixed borrowing limit, in the optimal contract investors can max out the borrowing limit, and the agent is always borrowing-constrained. The critical issue that a borrowing technology brings on is the agent's option to default. Without complication of default, the framework with CARA preference (with monetary effort cost) with borrowing and negative consumption allows for a tractable solution with private savings (see Williams 2006; He 2011). With default, the key is whether or not banks can seize the agent's private savings when he defaults. Bizer and DeMarzo (1999) point out that if banks can seize the agent's private savings in the default stage, then the debt-overhang problem (so the agent's marginal dollar of saving might go to banks when the personal debt is underwater) will discourage the agent's saving motive and restore the efficiency.

Agent's problem. Denote the agent's savings account balance by $S_t \geq 0$ (recall the borrowing constraint), which earns interest at the rate r. Given the contract Π, the agent's problem is[15]

$$\max_{\{a\},\{\widehat{c}\},\widehat{c}_\tau} \mathbb{E}^a \left[\int_0^\tau e^{-rt} \left[u\left(\widehat{c}_t\right) - \frac{b}{p}\left(a_t - p\right) \right] dt + e^{-r\tau} \frac{u\left(\widehat{c}_\tau\right)}{r} \right] \tag{4}$$
$$s.t.\ dS_t = rS_t dt + c_t dt - \widehat{c}_t dt \text{ with } S_0 = 0,\ S_t \geq 0 \text{ for } 0 \leq t < \tau,$$
$$\widehat{c}_\tau = r\left(F_\tau + S_\tau\right).$$

In Equation (4), $\mathbb{E}^a[\cdot]$ indicates that the probability measure is induced by the agent's effort policy $\{a\}$. $\{\widehat{c}\}$ and $\widehat{c}_\tau$ are privately observed consumptions during and after the employment, respectively. Note that the concavity of u implies that it is optimal for the agent to maintain a constant consumption level $\widehat{c}_\tau$ in his post-firing life. Therefore, the last term in Equation (4) $e^{-r\tau}\frac{u(\widehat{c}_\tau)}{r}$ captures the agent's value at the termination discounted back to time 0.

Investors' problem. We focus on the contract that implements working all the time so that $\{a_t = p\}$. The following lemma is a standard result in dynamic contracting (e.g., Cole and Kocherlakota 2001). Intuitively, whenever the agent wants to save, investors can save for him.

Lemma 1. Without loss of generality, we consider only contracts that induce no savings—that is, the agent always consumes his cash compensation.

We call the contract Π incentive-compatible and no-savings if $\{\{a\} = \{p\}, \{\widehat{c}\} = \{c\}, \widehat{c}_\tau = rF_\tau\}$ solves the problem in (4). The optimal contract solves the investors' problem:

$$\max_{\Pi \text{ is incentive-compatible and no-savings}} \mathbb{E}\left[\int_0^\tau e^{-rt}\left(YdN_t - c_t\right)dt - e^{-r\tau}F_\tau\right], \tag{5}$$

where $\mathbb{E}[\cdot]$ is under the probability measure induced by $\{a_t = p : 0 \leq t < \tau\}$—that is, the agent is working all the time before termination. Because the agent enjoys some nonnegative rents, in this problem the agent's time-0 participation constraint never binds. Denote by $\Pi^* = \left\{\{c^*\}, F^*_{\tau^*}, \tau^*\right\}$ the solution to this problem.

Remark 2. Cash compensation contract versus consumption contract. As is standard in this literature, the optimal contract characterizes only the optimal

---

[15] Heuristically, the sequence of events during $(t-dt,t]$ is: (i) the agent makes his effort decision a_t; (ii) the cash-flow realization (or not) is observed; (iii) the agent receives compensation c_t according to his performance; and (iv) the agent makes consumption/saving decision by choosing consumption $\widehat{c}_t$. This sequence ensures that the effort process is $\mathcal{F}^N$-predictable (i.e., does not depend on the cash-flow realization at $(t-dt,t]$), while the compensation process $\{c_t\}$ and the consumption process $\{\widehat{c}_t\}$ are $\mathcal{F}^N$-adapted (i.e., they can depend on the cash-flow realization at $(t-dt,t]$).

"*consumption*" policy that specifies how much the agent should consume given his performance history. Lemma 1 says that we are focusing on the contracting space where the derived *cash compensation* policy coincides with the agent's consumption policy. In general, the cash compensation contract that stipulates the history-dependent cash payment to the agent is just an implementation of the optimal consumption contract, and there might exist multiple compensation contracts that implement the same optimal compensation policy. For instance, there could exist another "optimal" cash compensation contract under which the agent follows the same optimal consumption policy $\{c\} = \{c^*\}$ while maintaining some positive savings by himself.

However, as we will discuss in Section 6.1.3, before reaching the absorbing first-best region (where the agent becomes risk-neutral with $u'\left(c_t^*\right) = \underline{\gamma}$), the agent faces a strict borrowing constraint under the optimal contract. As a result, investors have to do all the savings for the agent—otherwise the agent will withdraw from his private savings account and consume strictly more than c_t^*. Thus, as a history-dependent cash payment policy, the cash compensation rule derived in this article is indeed (essentially) the unique implementation of the optimal consumption contract.[16] For related discussions, as well as various implementations for the optimal cash compensation policy, see Section 6.1.3.

3. State Variables in the Relaxed Problem

We employ a relaxation method in this article. We first analyze two state variables that help us solve the relaxed problem. The first variable is the agent's continuation payoff, and the second one is the agent's marginal utility from consumption. Based on the agent's (local) joint deviation strategy, in Section 3.3 we specify the necessary conditions for the evolutions of two state variables and formulate the relaxed problem recursively with these necessary conditions only. We then solve the relaxed problem in Section 4, and Section 5 further verifies that the obtained solution is indeed the solution to the original problem stated in Equation (5).

3.1 Continuation Payoff and Incentive-compatibility Constraint

In this article, the term "incentive-compatibility constraint" is used exclusively for the agent's effort choice. In other words, we say that at any time t the contract is incentive-compatible, if the agent's single effort deviation—that is, from the equilibrium working effort $a_t = p$ to shirking $a_t = 0$ or myopic action $a_t = \overline{p}$—while fixing the follow-up effort-consumption policies cannot improve the agent's value.

[16] To be precise, the implementation is unique before the contract reaches the absorbing first-best region where the risk-neutral agent (with marginal utility $\underline{\gamma}$) holds a substantial stake in the firm. See a more detailed discussion in Section 6.1.3.

Given a contract $\Pi = \{\tau, \{c\}, F_\tau\}$, we introduce the agent's continuation payoff, W_t, as

$$W_t \equiv \mathbb{E}_t \left[\int_t^\tau e^{-r(s-t)} u\left(c_s\right) ds + e^{-r(\tau-t)} \frac{u\left(rF_\tau\right)}{r} \right]. \tag{6}$$

In other words, standing at time t, W_t is the agent's future value from the contract Π if he keeps working from then until termination, and conducts no savings. It is important to note that, in equilibrium, W_t has to be the agent's optimal value among those under other feasible deviation strategies.

Using Equation (6), the martingale representation result allows us to write the evolution of W as[17]

$$dW_t = rW_{t^-}dt - u\left(c_{t^-}\right)dt + \beta_t^W \left(dN_t - pdt\right), \tag{7}$$

where $\left\{\beta_t^W\right\}$ is some $\mathcal{F}^N$-predictable process. Economically, the martingale loading β_t^W measures the responsiveness of the agent's continuation payoff to the unexpected performance $dN_t - pdt$ under the equilibrium working effort. Fixing the agent's equilibrium consumption plans as recommended by the contract, it is β_t^W that controls the agent's effort choice. Intuitively, the agent makes his effort decision as follows. Choosing a_t affects the agent's effort cost $b\left(1 - \frac{a_t}{p}\right)dt$. However, this also sets the drift of dN_t to be a_t in his continuation payoff. As a result, the agent is solving

$$\max_{a_t \in \{0, p, \overline{p}\}} \; b\left(1 - \frac{a_t}{p}\right) + \beta_t^W a_t.$$

Since the objective is linear in a_t, β_t^W has to be equal to $\frac{b}{p}$ in order to implement the interior effort $a_t = p$.

Under the framework of binary effort levels (e.g., DeMarzo and Sannikov 2006; He 2009), to motivate working against shirking, the incentive β_t^W must be *no less than* $\frac{b}{p}$. Because the same argument can be applied to the effort choice between "working" and "myopic action," to prevent the agent from taking $a = \overline{p}$, β_t^W must be *no greater than* $\frac{b}{p}$. In other words, because highly powered incentives can induce some myopic actions from the agent, investors never impose excessive incentives on the agent. As a result, $\beta_t^W = \frac{b}{p}$. We have the following proposition, in line with Sannikov (2008).

Proposition 1. For any employment contract Π to be incentive-compatible, the agent's continuation payoff W_t evolves according to Equation (7), and $\beta_t^W = \frac{b}{p}$ for all $t \in [0, \tau)$ a.e.. This implies that the agent is indifferent between working and shirking—that is, he obtains the same value by taking $a_t = 0$ or p for any $t \in [0, \tau)$.

[17] See the proof of Proposition 1. Throughout the article, for processes involving jumps, dA_t is defined as $A_t - A_{t-dt}$ due to the right-continuous-left-limit (RCLL) property of the standard Poisson process.

For illustration, consider the following discrete-time example, which we will use again in the next subsection. Ignore discounting ($r = 0$), and set $p = 0.5$, $b = 2$. Suppose that at date t before consumption, the agent is promised with a continuation payoff of 11. Consider a contract where the agent's date t consumption $c_t = 1$, and assume that $u(1) = 1$. Then his postconsumption continuation payoff at date t is 10. In equilibrium, for promise keeping we must have

$$0.5 \times W^1_{t+1} + 0.5 \times W^0_{t+1} = 10, \tag{8}$$

where W^1_{t+1} (or W^0_{t+1}) is the preconsumption continuation payoff at date $t+1$ with (or without) success along the equilibrium path. Now it is clear that the reward difference $W^1_{t+1} - W^0_{t+1}$ pins down the agent's working incentives. To implement interior working, however, it must be true that $W^1_{t+1} = 12$ and $W^0_{t+1} = 8$. If not, say $W^1_{t+1} = 13$ (11) and $W^0_{t+1} = 7$ (9), and then the agent will take the myopic (shirking) action.[18] Here, the incentive loading $\beta^W_t = W^1_{t+1} - W^0_{t+1} = \frac{b}{p} = 4$. Note that if the agent shirks, his deviation payoff is $b + W^0_{t+1} = 10$, which is just his date-t postconsumption payoff under working along the equilibrium path.

3.2 Marginal utility

Now we investigate the agent's saving incentives. Denote by $M_t \equiv u'(c_t)$ the agent's marginal utility at time t. We have the following proposition, based on the requirement that working and not saving have to be optimal among all possible deviation strategies (therefore, W_t is indeed the optimal value that the agent can achieve from the continuation contract).

Proposition 2. The necessary conditions for Π to be incentive-compatible and no-savings are:

1. The continuation payoff process $\{W\}$ evolves according to Equation (7), where $\beta^W_t = \frac{b}{p}$ for all $t \in [0, \tau)$ a.e.;

2. For $0 \leq t < t' < \tau$, the agent's marginal utility process $\{M\}$ satisfies $\mathbb{E}^a_t[M_{t'}] \leq M_t$ a.e., where the agent's action $a_s = 0$ or p for $s \in (t, t')$.

To gain some intuition, we discuss the implications of the second condition regarding the equilibrium dynamics of M. To rule out private savings, the agent's expected marginal utility must be non-increasing over time (i.e., supermartingale). Otherwise, the agent can smooth his consumption and in

[18] For instance, suppose $W^1_{t+1} = 11$ and $W^0_{t+1} = 9$, then the equilibrium value from working is $0.5W^1_{t+1} + 0.5W^0_{t+1} = 10$. However, by shirking, the agent's deviation value is $b + W^0_{t+1} = 2 + 9 = 11 > 10$.

turn obtain a strictly higher deviation value. Heuristically, the marginal utility must satisfy

$$\mathbb{E}_{t-}^{a_t}\left[M_t\right] \le M_{t-}, \tag{9}$$

where $\mathbb{E}_{t-}^{a_t}\left[M_t\right]$ is the conditional expectation of M_t given effort choice a_t before knowing whether or not there is a cash flow in $(t-dt, t]$.

As a salient feature of any dynamic agency problem, the probability measure is induced by the agent's endogenous effort choice a_t. Then, under the equilibrium working effort, condition (9) requires that

$$\mathbb{E}_{t-}^{a_t=p}\left[M_t\right] = (1-pdt)\cdot M_t^0 + pM_t^1 dt \le M_{t-}, \tag{10}$$

where we denote by M_t^0 (M_t^1) the agent's marginal utility at t without (with) success during the interval $(t-dt, t]$.

More importantly, because the agent loses nothing from shirking (recall Proposition 1), the same result must hold for the off-equilibrium shirking effort $a_t = 0$. Specifically, when the agent shirks—so for sure there is no jump—condition (9) requires that

$$\mathbb{E}_{t-}^{a_t=0}\left[M_t\right] = M_t^0 \le M_{t-}. \tag{11}$$

This immediately implies a surprising result that on the path of no success the agent's marginal utility cannot rise. In other words, the optimal contract cannot cut the agent's consumption following his failures.

This result is based on the agent's potential joint deviation of shirking and saving. Following the previous discrete-time example discussed in Section 3.1, let us assume further that $u'(1) = 1$, $u'(0.8) = 1.1$, and $u'(1.2) = 0.9$. Recall that $c_t = 1$. Consider a hypothetical contract that assigns a lower consumption after poor performance—that is, set $c_{t+1}^0 = 0.8$ and $c_{t+1}^1 = 1.2$. Since $p = 0.5$ in this example, this contract satisfies the no-savings condition (10) under the measure induced by working. However, it violates the no-savings condition (11) under the measure induced by shirking, which opens the door for the following profitable joint deviation. Recall that at the end of Section 3.1, we have shown that, by deviating from working to shirking—but without changing the consumption/saving policy—the agent's preconsumption deviation continuation payoff at t remains at his equilibrium continuation payoff 11; that is,

$$11 = b + u(1) + u(0.8) + \left[W_{t+1}^0 - u(0.8)\right].$$

Here, with shirking, the agent's utility flow at date t is $b + u(1)$, and his preconsumption continuation value at $t+1$ is the sum of his consumption utility $u(0.8)$ and his postconsumption value $W_{t+1}^0 - u(0.8)$ (simply assume that the agent follows equilibrium strategies from date $t+1$ on). Now, if the agent saves

0.1 at date t and consumes this saving at date $t+1$, then his preconsumption joint-deviation value at t becomes

$$b+u\left(0.9\right)+u\left(0.9\right)+\left[W_{t+1}^{0}-u\left(0.8\right)\right]>b+u\left(1\right)+u\left(0.8\right)$$
$$+\left[W_{t+1}^{0}-u\left(0.8\right)\right]=11.$$

Therefore, this hypothetical contract fails to be incentive-compatible and no-savings.

Condition (11) states only that consumption cannot fall after failures. Using the dynamic programming approach, Section 5 rules out the possibility of lowering consumption after successes. With these two results, the optimal contract features the following:

$$M_t^1 \le M_t^0 = M_{t-}.$$

In words, in our optimal contract the agent's consumption (which is also his compensation paid by investors) is downward-rigid—it remains constant without jumps, but might rise in response to successes. However, as emphasized in the following remark, the important lesson from our analysis is that the private-saving consideration in general implies a greater downward rigidity in agent's compensation, which resembles the asymmetric payoff pattern of options. Empirically, this downward rigidity may be reflected in the wedge between incentives due to positive performance and incentives following poor performance, and in Section 6.2 we will discuss this result in greater detail.

Remark 3. Condition (11) and its implied perfect downward rigidity rely on the simplifying assumption that the probability of success under shirking is 0. If the probability of success under shirking is strictly positive, say $\varepsilon > 0$, then condition (11) under shirking becomes

$$(1-\varepsilon dt)\,M_t^0+\varepsilon M_t^1 dt \le M_{t-},$$

which implies that

$$M_t^0 - M_{t-} \le \varepsilon\left(M_t^0 - M_t^1\right)dt. \qquad (12)$$

In other words, in the optimal contract the marginal utility (consumption) could have a positive (negative) drift along the path of no success. Having said that, the downward punishment speed depends on the off-equilibrium measure implied by shirking, and Equation (12) shows that the positive drift of M (or the negative drift of c) after failures vanishes as ε gets close to zero.

Remark 4. Both the linearity of effort cost and the presence of myopic action play important roles in the analysis. Essentially, they force the incentive-compatibility constraint to be binding when the agent chooses working against shirking (Proposition 1). Without them, the contract may impose highly powered incentives $\beta_t^W > \frac{b}{p}$, and the agent loses $\beta_t^W - \frac{b}{p} > 0$ when he deviates to shirking. Because the agent finds himself strictly worse off by shirking, the local shirking-saving strategy illustrated by the previous numerical example might not be profitable,[19] and as a result we no longer have the key condition (11). See also the related discussions in Section 7.2 and Section A.8.2.

3.3 Formulating the relaxed problem

We have derived the necessary conditions in Proposition 2 for the contract to be incentive-compatible and no-savings. The relaxed problem just replaces the original constraints in Problem (5) with these necessary conditions. We rule out randomization (based on certain exogenous public signals) in solving the relaxed problem; this treatment is without loss of generality, as we will show shortly that the investors' value function without randomization is concave.

3.3.1 Dynamics of state variables. To be in line with the analysis of jump processes (e.g., Protter 1990; Biais et al. 2007), we use the left-hand limit of $\{W\}$ and $\{M\}$—that is, $W_{t-} \equiv \lim_{s\uparrow t} W_t$ and $M_{t-} \equiv \lim_{s\uparrow t} M_t$, as the state variables.[20] According to Proposition 2, the (left limit of) agent's continuation payoff W_{t-} follows:

$$dW_t = rW_{t-}dt - u(c_{t-})\,dt + \frac{b}{p}\,(dN_t - pdt)\,. \tag{13}$$

The agent's marginal utility serves as the second state variable in this model.[21] The following lemma gives a formal statement of the dynamics of M_{t-}. Here, $dM_t^D \leq 0$ in Equation (15) corresponds to condition (11); that is, there is no consumption cut after failures. And, $dM_t^D \leq -\beta_t^M pdt$ in Equation (16) corresponds to condition (10); that is, the marginal utility follows a supermartingale under the equilibrium measure induced by always working.

---

19 Note 23 in Sannikov (2008) gives an intuitive argument why a binding (local) incentive-compatibility constraint in the binary-effort setting (shirk or work) induces the agent to save privately. Mitchell and Zhang (2007) formally show that with binary effort levels the optimal contract features a slack local incentive-compatibility constraint; that is, $\beta_t^W > \frac{b}{p}$.

20 $J(W_t, M_t)$ and $J\left(W_{t-}, M_{t-}\right)$ differ only at (countably many, almost surely) points where a cash flow occurs, which is a zero measure set.

21 At first sight it seems that we can equivalently choose the nondecreasing c as the second state variable. However, to rule out randomization in the optimal contract, the marginal utility becomes the key variable in preventing the agent's private savings. We will formally show the concavity of the investors' value function (with arguments W and M) in Proposition 4 in Section 4.4, which implies that randomization is suboptimal.

Lemma 2. Condition 2 in Proposition 2 holds if and only if there exist two $\mathcal{F}^N$-predictable processes $\{\beta_t^M\}$ and $\{M_t^D\}$ for $t \in [0, \tau)$ such that[22]

$$dM_t = dM_t^D + \beta_t^M dN_t, \tag{14}$$

where

$$dM_t^D \leq 0, \text{ and} \tag{15}$$

$$dM_t^D \leq -\beta_t^M p dt. \tag{16}$$

3.3.2 Recursive formulation. Recall that in our model the agent can generate at most K cash flows. Let $W_{0-} \equiv W_0$ and $M_{0-} \equiv M_0$. Denote by $J^K(W_{0-}, M_{0-})$ the investors' value given the initial state variables W_{0-} and M_{0-}, where K denotes the number of remaining cash flows. The relaxed problem, in its recursive formulation, is

$$J^K(W_{0-}, M_{0-}) = \max \mathbb{E}\left[\int_0^\tau e^{-rt}(YdN_t - c_t)\,dt - e^{-r\tau}F_\tau\right],$$

subject to constraints (13), (14), (15), and (16).

4. Solution to the Relaxed Problem

Using the dynamic programming technique, we solve the relaxed problem in this section in a heuristic way. Section 5 formally verifies that the solution solves the relaxed problem.

4.1 Preliminaries

We can construct the investors' value function $J^K(W_{t-}, M_{t-})$ iteratively (see Appendix A.6). Because the key properties of the value function are independent of K, for illustrative purposes in the main text we take K to infinity, and define $J(W_{t-}, M_{t-}) \equiv J^\infty(W_{t-}, M_{t-})$. For simpler notation, we suppress the subscript t^- most of the time so that (W, M) corresponds to (W_{t-}, M_{t-}).

Several functions are useful in later analysis. It is clear that the agent's marginal utility $M \in [\gamma_L, \gamma]$. Since the analysis is trivial for $M = \gamma_L$ (i.e., the agent becomes risk-neutral; see Section 4.5 for this case), we focus on the strictly concave part of the agent's utility function in Equation (3). To express the agent's utility and consumption in terms of marginal utility M, we define the utility function as

$$U(M) \equiv 1 - \frac{M}{\gamma}, \tag{17}$$

[22] We focus on the employment path $t \in [0, \tau)$. After the agent is fired at τ, there are no further cash flows, and consumption smoothing implies that $dM_t = dM_t^D = 0$.

and the consumption function as

$$c(M) \equiv \frac{1}{\gamma} \ln \frac{\gamma}{M}. \tag{18}$$

When the agent is fired, to fulfill the continuation payoff W investors simply pay the agent a lump-sum transfer of $F_\tau = \frac{u^{-1}(rW)}{r}$. Therefore, we define the investors' value function at termination as (recall that we have normalized the firm's liquidation value $L = 0$)

$$J^L(W) \equiv \frac{u^{-1}(rW)}{r}. \tag{19}$$

4.2 Optimal contract and the timeline

To solve the relaxed problem, we take a guess-and-verify approach. The first step is to guess the optimal policy, as illustrated in the timeline in Figure 2. It highlights the subperiod for the n^{th} cash flow. As shown, we decompose

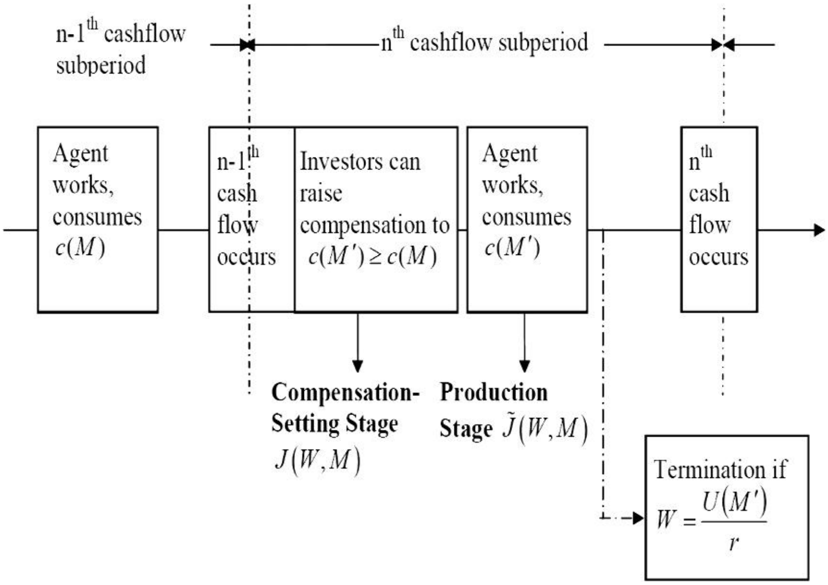

Figure 2
Timeline of optimal contracting
The n^{th} cash-flow subperiod starts with the occurrence of the $(n-1)^{th}$ cash flow. Investors can raise the agent's compensation (compensation-setting stage with value function $J(W,M)$) from $c(M)$ to $c(M')$. Afterward the agent works to produce the n^{th} cash flow (production stage with value function $\widetilde{J}(W,M)$). The project is liquidated and the agent is fired if his continuation payoff W hits $\frac{U(M')}{r}$ before the occurrence of the n^{th} cash flow. These two stages repeat themselves for the following subperiods.

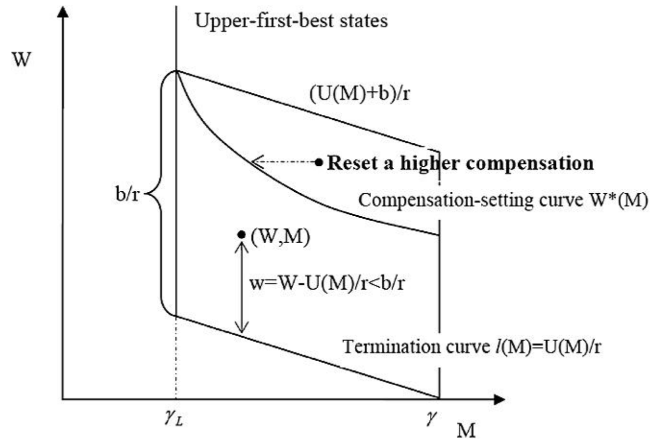

Figure 3
The (W, M) state space
The liquidation line is $l(M) = \frac{U(M)}{r}$, and the compensation-setting curve $W^*(M) < \frac{U(M)+b}{r}$ is downward sloping; that is, $W^{*\prime}(M) < 0$. Whenever (W, M) is above the curve $W^*(M)$, the optimal contract resets a higher pay level $c(M')$; that is, a lower marginal utility M' so that $W = W^*(M')$.

each subperiod into the *compensation-setting* stage and the *production* stage. Given an occurrence of cash flow, in the compensation-setting stage investors have the option to raise the agent's compensation from $c(M)$ to $c(M')$, which corresponds to the marginal utility response β_t^M in Equation (14). Then we enter the production stage, in which the agent keeps working $(a_t = p)$ until the n^{th} cash flow realizes, or is fired before the n^{th} cash flow realization.

As shown in Figure 3, $J(W, M)$ as the value function of the compensation-setting stage incorporates the investors' option value to raise the agent's compensation before they ask the agent to work. Denote by $\widetilde{J}(W, M)$ the value function of the production stage, which excludes this option value.

4.3 Production Stage: Construction of $\widetilde{J}(W, M)$

The construction is backward. Specifically, we first take as given the value function $J(W, M)$ in the compensation-setting stage after the n^{th} cash flow. Then we move backward to consider the production and compensation-setting stages in the n^{th} cash-flow subperiod, which is the time interval after the $(n-1)^{th}$ cash-flow but before the n^{th} one.

As shown in Figure 2, the $(n-1)^{th}$ cash flow occurs at t_{n-1}, and both parties enter the compensation-setting stage. Suppose that in the compensation-setting stage investors set a marginal utility $M_{t_{n-1}}$ for the agent who has been promised by a continuation payoff $W_{t_{n-1}}$. Then the pair $(W_{t_{n-1}}, M_{t_{n-1}})$ sets the initial state of the production stage.

4.3.1 Dynamics of state variables. Due to Equation (13), without successes W evolves according to

$$dW_t = rW_{t^-}dt - U(M_{t^-})\,dt - bdt. \tag{20}$$

Recall that $dM_t = dM_t^D + \beta_t^M dN_t$ in Equation (14). We verify shortly that in the production stage it is optimal to set

$$dM_t^D = 0. \tag{21}$$

As a result, the agent's marginal utility $M_{t^-} = M$ remains constant without successes. Once a cash flow occurs, the agent's continuation payoff jumps to $W_{t^-} + \frac{b}{p}$, and investors obtain a value $J\left(W_{t^-} + \frac{b}{p}, M\right)$.[23]

4.3.2 Termination. In our model, one form of inefficiency comes from early termination/firing. Define by $l(M) \equiv \frac{U(M)}{r}$ the termination line as shown in Figure 3. The following lemma characterizes the termination/firing.

Lemma 3. When $W = l(M) = \frac{U(M)}{r}$, the agent is fired and the firm is liquidated.

This result can be understood as follows. Due to potential shirking benefit, $W - \frac{U(M)}{r}$ reflects the positive rent enjoyed by the agent. When $W = \frac{U(M)}{r}$, zero future rent triggers an immediate dismissal. Here, although the agent is fired due to his poor performance, he is granted a total transfer $F = \frac{c(M)}{r}$ in this "punishing" termination event, which corresponds to severance pay in implementation. Therefore, we have (recall Equation (19))

$$\widetilde{J}\left(\frac{U(M)}{r}, M\right) = J^L\left(\frac{U(M)}{r}\right) = -\frac{c(M)}{r}.$$

4.3.3 Value function $\widetilde{J}(W, M)$ and its properties. In the region $W > \frac{U(M)}{r}$, Equation (13) and constant M before any success imply a Hamilton-Jacobi-Bellman (HJB) equation for the investors' value function $\widetilde{J}$:

$$r\widetilde{J}(W,M) = pY - c(M) + p\left[J\left(W + \frac{b}{p}, M\right) - \widetilde{J}(W,M)\right] + \widetilde{J}_W(W,M)\left(rW - U(M) - b\right). \tag{22}$$

The left-hand side is the investors' required return. On the right-hand side, the first term is the expected cash flow, and the second term is the compensation

[23] Although there is no corresponding jump β_t^M in M here, keep in mind that $J(\cdot,\cdot)$, as the value function of the compensation-setting stage, has taken into account the option of reducing M in response to a cash flow. Section 4.4 studies the optimal response of β_t^M given a cash flow in the compensation-setting stage.

payment. The last two terms capture the value change due to the evolution of state variable W: the third term is the expected value change due to the jump from W to $W + \frac{b}{p}$, and the fourth term is the value change due to the drift of W without jump.

In the next subsection, we will show that $W < \frac{b+U(M)}{r}$ along the equilibrium path. Intuitively, the perpetuity of consumption utility plus shirking benefit, that is, $\frac{U(M)+b}{r}$, is the highest value that investors can possibly deliver to the agent given the pay level $c(M)$.

Given $W < \frac{b+U(M)}{r}$, the ordinary differential equation (ODE) in Equation (22) admits a closed-form solution:

$$\widetilde{J}(W, M) = [b - rW + U(M)]^{1+\frac{p}{r}} \left[\int_{\frac{U(M)}{r}}^{W} \frac{pY + pJ\left(x + \frac{b}{p}, M\right) - c(M)}{[b - rx + U(M)]^{2+\frac{p}{r}}} \times dx + J^L\left(\frac{U(M)}{r}\right) b^{-1-\frac{p}{r}} \right]. \tag{23}$$

One can read the solution as follows: at any state (W', M), investors' instantaneous net gain is simply

$$\left(Y + J\left(W' + \frac{b}{p}, M\right)\right) \cdot pdt - c(M)\, dt,$$

which is the expected value upon success (with probability pdt) minus the outflow of compensation payment. Therefore, the investors' value at state (W, M) is the integration over these instantaneous net gains for $W' < W$, plus the liquidation value $J^L\left(\frac{U(M)}{r}\right)$ in the scenario where the agent is fired before he delivers any cash flow. In Equation (23), these two sources of value are properly weighted according to the Poisson structure.

We list the main properties of the production stage value function $\widetilde{J}$ in Proposition 3. As the fixed-point argument suggests, they are based on the properties of J in the compensation-setting stage, which we study in the next subsection.

Proposition 3. For the production stage, the value function $\widetilde{J}(W, M)$ satisfies:

1. $\widetilde{J}_W \geq -\frac{1}{\gamma_L}$, and $\frac{1}{\gamma r M} < \widetilde{J}_M - \frac{1}{\gamma r}\widetilde{J}_W \leq \frac{1}{\gamma r \gamma_L}$.
2. $\widetilde{J}_{WW} < 0$, $\widetilde{J}_{MM} < 0$, and $\widetilde{J}_{WW}\widetilde{J}_{MM} - \left(\widetilde{J}_{WM}\right)^2 > 0$. Therefore, $\widetilde{J}(W, M)$ is concave.
3. $\widetilde{J}_{WM} < 0$, and $\widetilde{J}_M\left(\frac{b+U(M)}{r}, M\right) < 0$.

For property 1, because it costs investors at most $\frac{1}{\gamma_L}$ to deliver one unit of continuation payoff W, $\widetilde{J}_W$ is bounded by $-\frac{1}{\gamma_L}$. And, as shown in Figure 3, the endogenous termination probability is determined by $\omega \equiv W - \frac{U(M)}{r} = W - \frac{1}{r} + \frac{M}{\gamma r}$. Therefore, $-\widetilde{J}_M + \frac{1}{\gamma r}\widetilde{J}_W$ measures the (negative) impact on investors' value by reducing M (raising the agent's compensation) while fixing the termination probability (keeping ω constant). Given this intuition, the estimation result $\frac{1}{\gamma r M} < \widetilde{J}_M - \frac{1}{\gamma r}\widetilde{J}_W \leq \frac{1}{\gamma r \gamma_L}$ follows from the fact that the pay raise has to be permanent.[24]

The concavity of $\widetilde{J}$ in property 2 implies that any randomization beyond cash flow shocks is suboptimal. To ensure concavity, we need the following sufficient condition on the project's profitability (which is used in the proof of Lemma 5 in Appendix A.6):

$$Y > \max\left(\frac{1}{\gamma r}\left[\frac{\gamma}{\gamma_L} - 1\right]^2, \frac{b}{p\gamma_L}\right). \tag{24}$$

The third property pertains to the optimal compensation-setting policy, which we will turn to in the next subsection.

4.4 Compensation-setting stage: Construction of $J(W, M)$

4.4.1 Possible pay raise and properties of $J(W, M)$. Recall that at time t_{n-1} the $(n-1)^{th}$ cash flow occurs. Suppose that $M_{t_{n-1}^-} = M$, and the agent now has a continuation payoff $W = W_{t_{n-1}^-} + \frac{b}{p}$. If investors decide to keep the same marginal utility (i.e., set $M_{t_{n-1}} = M = M_{t_{n-1}^-}$) and enter the production stage, then they get a value $\widetilde{J}(W, M)$ as shown in the previous section. However, investors have the option to raise the agent's compensation (or reduce M) and enter the production stage with a new state pair (W, M'). Of course, this option is valuable only if investors can find $M' < M$ so that $\widetilde{J}(W, M') > \widetilde{J}(W, M)$.

Following this idea, we define the optimal marginal utility level M^*, as a function of W, as

$$M^*(W) \equiv \underset{M' \in [\gamma_L, \gamma]}{\arg\max}\ \widetilde{J}(W, M'). \tag{25}$$

Define investors' value function at the compensation-setting stage as

$$J(W, M) \equiv \begin{cases} \widetilde{J}(W, M) & \text{if } M \leq M^*(W) \\ \widetilde{J}(W, M^*(W)) & \text{otherwise} \end{cases}. \tag{26}$$

[24] To keep ω constant, a unit decrease in M has to be accompanied by a $\frac{1}{\gamma r}$-unit increase in W. This explains $-\widetilde{J}_M + \frac{1}{\gamma r}\widetilde{J}_W$. Also, because the future marginal utility $M_s \leq M_t = M$ where $s > t$, the marginal cost of permanently reducing one unit of M is a weighted average of $-\frac{c'(M_s)}{r}$ in the future, which must belong to $\left(\frac{1}{\gamma r M}, \frac{1}{\gamma r \gamma_L}\right]$.

Simply put, whenever the realization of cash flow brings state (W, M) above the curve $M^*(W)$, investors reduce M to $M^*(W)$ by exercising the option of raising the agent's compensation, as shown in Figure 3. Therefore, the optimal response of marginal utility M to a cash-flow realization at t is

$$\beta_t^M = \min\left(M^*\left(W_{t^-} + \frac{b}{p}\right) - M_{t^-}, 0\right). \tag{27}$$

The transformation in Equation (26) implies $J_M \geq 0$ always. Furthermore, function J inherits the concavity property from function $\widetilde{J}$. The following proposition gives the properties of $J(W, M)$ based on Proposition 3.

Proposition 4. For the compensation-setting stage, the value function $J(W, M)$ satisfies:

1. $J_W \geq -\frac{1}{\gamma_L}$, and $\frac{1}{\gamma r M} < J_M - \frac{1}{\gamma r} J_W \leq \frac{1}{\gamma r \gamma_L}$.
2. $J_{WW} < 0$, $J_{MM} \leq 0$, and $J_{WW} J_{MM} - (J_{WM})^2 \geq 0$. Therefore, $J(W, M)$ is concave.
3. $J_{WM} \leq 0$; $J_M \geq 0$ and $J_M\left(\frac{b+U(M)}{r}, M\right) = 0$.

4.4.2 Trade-off of compensation-setting. The economic rationale behind the compensation-setting policy is the trade-off between the termination cost and the consumption smoothing benefit. On the cost side, as $w = W - \frac{U(M)}{r}$ captures the distance to liquidation (see Figure 3), a smaller M (therefore a higher c) reduces w, leading to a higher termination probability. Intuitively, given a promised continuation payoff W, the agent's future rent (beyond his compensation guarantee) will be smaller for a higher c. This implies a more stringent punishment scheme, which makes the costly termination more likely. On the benefit side, due to the agent's risk aversion, raising compensation gives a consumption-smoothing benefit (as the agent's equilibrium consumption pattern is back-loaded). Consequently, the optimal compensation-setting policy equates the marginal cost (from inefficient terminations) with the marginal benefit (from consumption smoothing).

This trade-off is reflected in property 3 in Proposition 3. First, $\widetilde{J}_{WM} < 0$ implies that for $W > W'$, $-\widetilde{J}_M(W, M) > -\widetilde{J}_M(W', M)$, where $-\widetilde{J}_M$ captures the marginal benefit of raising c. In words, a higher continuation payoff W mandates investors to pay more in the future, leading to a higher consumption-smoothing benefit.

Second, $\widetilde{J}_M\left(\frac{b+U(M)}{r}, M\right) < 0$ implies that the curve $M^*(W)$ stays below the line $W = \frac{b+U(M)}{r}$. Put differently, it is always optimal to set a higher pay when $W = \frac{b+U(M)}{r}$. In Section 4.3.3, we note that $\frac{b+U(M)}{r}$ is the upper bound of the agent's continuation payoff W given M, and a pay raise is necessary following a success. Therefore, at $W = \frac{b+U(M)}{r}$, the marginal benefit of consumption smoothing is strictly positive. From the cost side, the marginal impact of future termination cost by setting $M^*(W)$ slightly below $\frac{b+U(M)}{r}$ is zero (see Appendix A.6.2). Thus, the benefit side dominates, and pay raise is optimal. Consistent with this logic, we have $W^*(\gamma_L) = \frac{b+U(\gamma_L)}{r}$ because of the zero consumption-smoothing benefit for a risk-neutral agent.[25]

Third, the definition of $M^*(W)$ in Equation (25) implies that $\widetilde{J}_M(W, M^*(W)) = 0$. The compensation-setting curve $M^*(W)$ is downward sloping, as shown in Figure 3, because we have

$$M^{*\prime}(W) = -\frac{\widetilde{J}_{WM}}{\widetilde{J}_{MM}} < 0.$$

This result allows us to define the inverse function $W^*(M)$, which is the highest continuation payoff given M such that $\widetilde{J}_M$ remains nonnegative.

4.4.3 Pay raises without success? So far we have ruled out raising compensation without successes. In other words, investors will not exercise the pay raise option along the path without successes (i.e., keep M constant). This is the implicit assumption $dM_t^D = 0$ in Equation (21) that we use in deriving $\widetilde{J}$ in Section 4.3.

To show this result, note that for states below the curve $M^*(W)$ we have $J_{WM} = \widetilde{J}_{WM} < 0$ due to the way that Equation (26) is constructed. After setting the compensation, the production stage must start from some state $\left(W_{t_{n-1}}, M\right)$ on or below $M^*(W)$, such that $J_M\left(W_{t_{n-1}}, M\right) \geq 0$ (see Figure 3). Along the path without any success, we have $W_{t-} < W_{t_{n-1}}$ for $(t \in t_{n-1}, t_n]$, where t_n is the time when the n^{th} cash flow occurs. But $J_{WM} < 0$ implies that $J_M(W_{t-}, M) > J_M\left(W_{t_{n-1}}, M\right) \geq 0$, and therefore pay raise (reducing M) is suboptimal along the path without any success. Intuitively, the marginal benefit of raising pay is smaller for subsequent lower continuation payoffs. If it is optimal to maintain the pay level when $W = W_{t_{n-1}}$ initially, then it must also be the case along the path without any success.

4.5 Upper-first-best Region

The above analysis does not cover the upper-first-best region $\left\{(W, M) : W \geq \frac{b+U(\gamma_L)}{r}, M = \gamma_L\right\}$ where the agent becomes risk-neutral; see Figure 3. In that region, the optimal contract is straightforward: the risk-neutral agent with $W \geq \frac{b+U(\gamma_L)}{r}$ consumes his compensation, which is never below $c(\gamma_L)$, keeps working always, and obtains $\frac{b}{p\gamma_L}$ from each cash-flow

25 To see this, according to property 1 in Proposition 3, when $M = \gamma_L$ we have $\widetilde{J}_M = \frac{1}{\gamma r}\left(\widetilde{J}_W + \frac{1}{\gamma_L}\right)$ always. Therefore, when $W = \frac{U(\gamma_L)+b}{r}$, the first-best result holds, and $\widetilde{J}_W = -\frac{1}{\gamma_L}$ implies $\widetilde{J}_M = 0$.

realization Y (recall the parameter constraint (24)). The first-best region is absorbing, and there is no future inefficient termination. For derivations of J in this upper-first-best region, see Appendix A.6.3.

5. Verification of the Optimal Contract

5.1 Verifying the optimal solution to the relaxed problem

We first verify that the contract described in Section 4 solves the relaxed problem formulated in Section 3.3. For details, see the proof for Proposition 5 in the Appendix.

Proposition 5. Consider the stationary case $K \to \infty$. The investors' value function $J(W, M) = J^{\infty}(W, M)$ exists, with properties established in Proposition 4, and the compensation-setting curve $M^*(W)$ defined in Equation (25) satisfies $M^{*\prime}(W) < 0$. Under the optimal solution to the relaxed problem formulated in Section 3.3, W_t evolves according to Equation (13) and M_t evolves according to Equation (14), where $dM_t^D = 0$ as in Equation (21) and $\beta_t^M = \min\left(M^*\left(W_{t-} + \frac{b}{p}\right) - M_{t-}, 0\right)$ as in Equation (27).

5.2 Verifying the optimal contract for the original problem

Now we show that the solution to the relaxed problem is also the solution to the original problem. The key observation is that, under the perfect downward-rigid compensation contract stated in Proposition 5, the agent's optimal strategy is always to exert working effort and maintain zero savings. In words, the obtained solution not only satisfies the necessary conditions identified in Proposition 2, but also satisfies the tighter constraints (i.e., a smaller set of feasible contracts) imposed by the original problem (5). As a result, the solution in Proposition 5 solves the investors' original problem (5). We have the following main theorem.

Theorem 5. Under the optimal contract Π^* that implements working, we have

$$dW_t = (rW_{t-} - U(M_{t-}) - b)\,dt + \frac{b}{p}dN_t,$$

and $dM_t^D = 0$, $\beta_t^M = \min\left(M^*\left(W_{t-} + \frac{b}{p}\right) - M_{t-}, 0\right)$ so that

$$dM_t = \beta_t^M dN_t.$$

The employment is started at the state $(W_0, M_0) = \arg\max_{(W,M)} J(W, M)$, and terminated whenever $W_\tau = \frac{U(M_\tau)}{r}$ so that the agent gets a transfer

$F_\tau = \frac{c(M_\tau)}{r}$.[26] Once $W_{t-} > \frac{U(\gamma_L)+b}{r}$, we have $M_{t-} = M^*(W_{t-}) = \gamma_L$, and the first-best result is achieved: Investors pay the agent $\frac{1}{\gamma_L}\left[W_{t-} - \frac{U(\gamma_L)+b}{r}\right]$, ask him to work forever, and pay him $\frac{b}{p\gamma_L}$ whenever a cash flow occurs.

Finally, we have to verify that when the loss Δ due to myopic action is sufficiently high, it is always optimal to implement working. For details, see Appendix A.8.2.

6. Discussions and Empirical Predications

In this section, we discuss implementation of the optimal contract, compare our results with those of Sannikov (2008) and Harris and Holmstrom (1982), and make an attempt to relate our theoretical results to the compensation contracts observed in practice.

6.1 Implementation and comparison to literature

One straightforward implementation of the optimal contract is as follows. In this employment contract, the agent is offered a lifetime wage guarantee. If the agent's performance is sufficiently good, he will receive wage raises (as promotions), and these raises are permanent. In contrast, given poor performance, the agent is dismissed with severance pay to support his post-firing consumption at his current wage level, and he loses potential future pay raises.

6.1.1 Comparison to Sannikov (2008) with contractible savings. The possibility of private savings has a dramatic impact on the optimal compensation policy. Let us make only one modification to our model and consider the case that the agent's savings are (publicly) observable and contractible. In other words, the agent's consumption (which is just the compensation paid by investors) is contractible. As a dynamic agency problem with hidden actions studied in Sannikov (2008), the agent's continuation payoff W_{t-} is the only state variable in solving for the optimal contract (see Appendix A.9).[27]

We graph the optimal compensation policies (the left scale) and associated continuation payoff dynamics (the right scale) in Figure 4. The history consists of four cash flows at $t = 0.5$, 1.0, 1.5, and 3.5; afterward, the agent generates no cash flows even with his effort input. The top (bottom) panel is for the case with observable (private) savings that we used in Figure 1. For better comparison, we use the same scale for both cases.

[26] Using τ or τ^- makes no difference, because termination cannot occur at the exact time point of cash-flow realization (given a success, the agent's continuation payoff $W_{t-} + \frac{b}{p} > \frac{U(M)}{r}$ given $W_{t-} \geq \frac{U(M)}{r}$).

[27] In the binary-effort version of Sannikov (2008), there are no myopic actions. But because in Sannikov (2008) the optimal contract features a binding incentive-compatibility constraint, the restriction brought on by myopic actions is redundant.

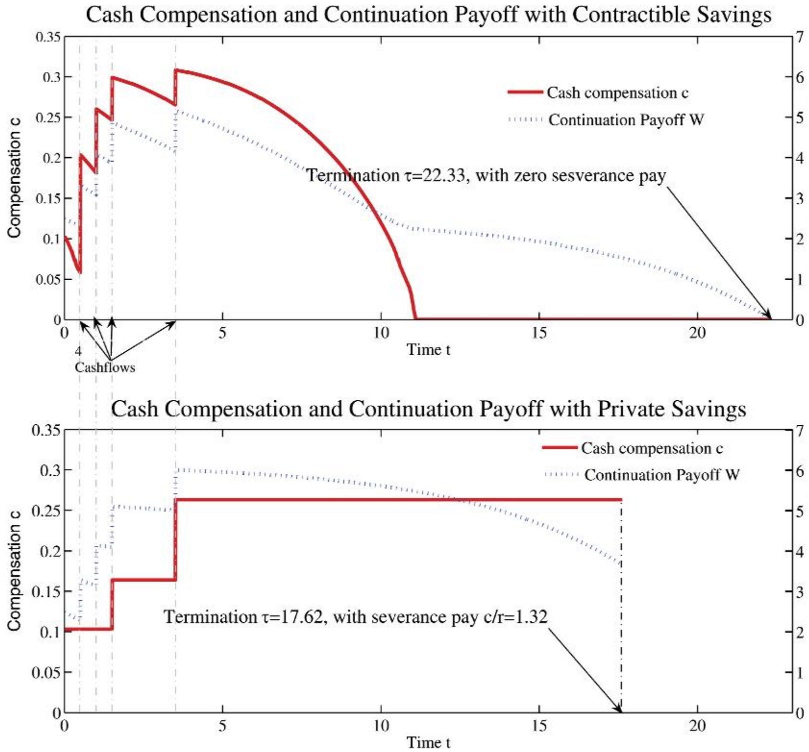

Figure 4
Optimal cash compensation policies and associated continuation payoff evolutions for the cases with private savings (the bottom panel) and the case with observable savings (the top panel)
The solid line is for cash compensation process, and the dotted line is for the agent's continuation payoff W. The history consists of four cash flows at $t = 0.5$, 1.0, 1.5, and 3.5, and no cash flow afterward. Parameters are $b = 0.5$, $Y = 20$, $r = 0.2$, $p = 0.5$, $\gamma = 5$, and $\gamma_L = 1$.

In the top panel with observable savings, the agent's pay exhibits a very sensitive response (a zig-zag pattern) to his performance: his compensation goes up for any success, and drops given no success. In contrast, in the bottom panel with private savings, the response is muted: compensation displays a downward-rigid pattern, and pay raises are less frequent (only twice given four cash flows). Put differently, the agent's pay might go up or stay the same following successes, but he never gets a pay cut after poor performance. This rigidity only to poor performance resembles the asymmetry pattern in options payoffs in executives' remuneration contracts, and we will come back to this issue in Section 6.2, where we discuss empirical predictions of our model.

It is interesting to stress that the downward rigidity of cash compensation does not suppress the agent's working incentives. In fact, in both panels, the agent faces the exact same incentives to exert effort. Due to the dynamic nature of long-term contracting, the agent's incentives depend only on how responsive his continuation payoff W is to his performance. In the bottom panel, despite the downward-rigid pay, the agent's continuation payoff goes down on the path without successes, because the agent slowly loses the chance for future promotions. Mapping to reality, the response of the agent's continuation payoff, without noticeable changes in his current cash compensation, corresponds to the change in the value of the restricted stocks held by corporate managers.

Finally, private savings also have remarkable impact on the termination policy. Given the long poor performance after $t = 3.5$ in Figure 4, the agent's continuation payoff falls at a lower rate in the top panel than in the bottom panel. This is because of the downward compensation adjustment along the path of poor performance when savings are contractible. As a result, with contractible savings (the top panel) the firm's life span is longer. Besides, in contrast to zero severance pay when savings are contractible, the agent in our model walks away with a positive severance pay. Economically, this harsher termination policy in forced turnovers is necessary for maintaining proper incentives, because the agent's cash compensation contract is relatively lenient. Empirical discussion in Section 6.2 will emphasize this predicted positive association between downward-rigid cash compensation and harsher termination policy, which can potentially distinguish our theory from the entrenchment theory proposed by Bebchuk and Fried (2004).

6.1.2 Discussions with Harris and Holmstrom (1982). Our model can be applied to any long-term labor compensation contracts. In fact, our particular form of implementation resembles the compensation contract received by relatively low-rank employees in certain industries. For instance, pilots in the airline industry earn hourly wages that increase with their ranks and have a certain significant amount of severance pay.[28]

Harris and Holmstrom (1982) also derive a downward-rigid wage contract to be the optimal contract. In that model, both the learning about the agent's ability and the firm's one-sided commitment are the driving forces. In contrast, we obtain the same dynamic structure for the optimal contract under a framework with moral hazard only. In this regard, the theoretical predictions from our model are consistent with the empirical evidence mentioned in Harris and Holmstrom (1982); that is, the positive relationship between experience and earnings, the positive skewness of earnings, and so on.

Even though Harris and Holmstrom (1982) and this article generate similar results, it is possible to separate these two theories empirically. Start with an agent who receives a pay raise, and focus on how the ordering of his follow-up performance (i.e., whether successes come before failures) affects his next pay

[28] For instance, Delta airline pilots receive up to nine months of severance pay, according to http://www.usatoday.com/travel/flights/2009-05-28-delta-pilot-retirement_N.htm. Of course, bear in mind that the specific contract form might be just a superficial similarity, rather than driven by the exact economic force analyzed in this article.

raise. In typical learning models such as Harris and Holmstrom's (1982), the agent's performance is his underlying ability plus some i.i.d. noises, and the ordering of performance does not matter. That is because the simple average of these performances is a sufficient statistic to update the agent's perceived ability, which determines the agent's pay raise, if any. In contrast, in our model, earlier successes lead to a higher continuation payoff W (due to the simple discounting effect $r > 0$; check Equation (13)), and as a result the following pay raise should be greater.[29]

6.1.3 Uniqueness of optimal cash compensation policy. In Remark 2 we emphasized that, in general, the optimal contract can only uniquely pin down the optimal *consumption* policy—that is, the optimal amount that the agent should consume given his performance history. It is natural to ask the following question: Is the optimal "cash compensation" policy unique among all potential implementations of the optimal consumption policy? By focusing on the "no-saving" cash compensation contract, essentially we restrict our attention to a class of cash compensation contracts in which investors do all the savings for the agent. Can investors take a different cash compensation scheme, in which the agent saves for himself (hence *truly private savings*) while still achieving the desired consumption policy?

The answer is essentially no. As we have shown, the key feature of our optimal consumption policy is that, before the contract reaches the first-best region $M = \gamma_L$, along the optimal path the agent is strictly borrowing-constrained. In other words, along the optimal path there will be some states in which the agent knows that his consumption will go up at the next instant once he delivers a success, and hence his expected marginal utility goes down. Now consider another implementation where the agent saves some cash in his private savings account. Then, at these states, the agent will engage in consumption smoothing to consume strictly more than the level stipulated by the optimal consumption policy, and this deviation cannot be punished in the optimal contract due to the private-saving assumption. As a result, in any implementation of the optimal consumption contract, the cash compensation pattern is indeed unique before the first-best region of $M = \gamma_L$ is reached.[30] Of course, from a more practical point of view, although the cash

[29] Consider the following two histories after the first pay raise: (1, 1, 0, 1) and (0, 1, 1, 1), where 1 (0) indicates success (failure), and suppose that there is a second pay raise after these four realizations. Our model predicts a greater pay raise in the first path with more successes in early times, while Harris and Holmstrom (1982) imply that we should observe the same pay raise for both paths (if the noise variance is time-varying so that performance ordering does not matter for learning). Of course, this argument relies on the assumption that the implemented effort is independent of the performance history (we thank an anonymous referee for pointing this out).

[30] Strictly speaking, the optimal contract can still ask the agent to hold some positive amount of cash at states with low continuation payoffs so that $\beta_t^M = 0$; that is, the agent's marginal utility does not go down even with a cash-flow occurrence at that moment. But eventually, after a *good* (not bad) performance history, $\beta_t^M < 0$ holds (see Figure 3), and the agent has to return these savings back to investors. This awkward implementation runs against the limited liability of the agent who can claim he has consumed it (though the cash is actually sitting

compensation policy is essentially unique, the detailed implementation of the cash compensation policy can be fairly flexible. For instance, other than annual salary and cash bonuses, corporate managers are also paid by restricted stocks and stock options that eventually pay out cash. And, the termination transfer can include severance pay as well as pension plans and vested stocks.

6.2 Empirical discussions and predictions

In this section, we provide empirical discussions based on our theory. Our modeling assumptions—although extremely stylized—capture certain important aspects in agency frictions between firms and their managers,[31] and we now discuss the model's implications on executive compensation.

6.2.1 General patterns that are consistent with the model. We emphasize the following general patterns in executive compensation that are consistent with our theory:

1. Consistent with the popularity of options-type remuneration contracts in practice, our model predicts that cash compensation will be rigid to the managers' poor performance. It is worthwhile to stress that the fact that managers receive performance-dependent cash compensation (including salary, bonus, options grants, etc.) is not directly in contradiction to our model. As we emphasized in Remark 3, with a positive probability of success under shirking—instead of zero as assumed in the analysis—the optimal cash compensation will indeed go down after poor performance. Therefore, the general message delivered by our model is that since the agent can potentially undo his on-the-job incentives, the resulting compensation contract tends to feature a greater rigidity (rather than the perfect downward rigidity) relative to those derived from models with contractible savings.

2. Consistent with Kaplan and Minton (2008) and Yermack (2006), our model predicts *forced turnovers but with sizable severance pay*, a new theoretical feature among dynamic agency models (e.g., Sannikov 2008). It is important to stress that firing in our model is a punishment to the agent. Despite the severance package upon firing, the agent loses his entire option value of being promoted in the future. This mechanism captures certain aspects in the real world, as managers who leave a

in his private account). In addition, it contradicts with the practice of "bonus clawback" where the agent has to return to investors some portion of his previously earned bonuses after *bad* performance.

[31] Among the assumptions made in this article, the only one that seems counterfactual is the borrowing constraint. Admittedly, in the real world, top CEOs do not seem to be borrowing-constrained, with the possible exception that some private-company CEOs borrow to buy equity in the firm (we thank an anonymous referee for this point). In the model, the ability to borrow essentially places an upward rigidity pressure—that is, a limit on the extent that the agent's marginal utility can fall (or consumption/compensation can rise)—and our main result still holds when there is a rate differential between borrowing and lending. See note 14.

firm involuntarily may suffer from their tainted reputation, in addition to losing their existing but not-yet-vested options in the firm (Rubinstein 1995).

3. A straightforward extension of our model shows that the size of a severance package decreases with the agent's outside option upon leaving the firm. Recall that in the base model the agent has a zero outside option. Now suppose that the agent receives a constant z in perpetuity after his layoff; the optimal contract will simply specify a severance package with $\max\left[c\left(M\right)-z,0\right]/r$ to prevent the agent's consumption from falling after the layoff,[32] which is decreasing in the outside option z. This prediction is consistent with certain contracts observed in practice, and keep in mind that the manager's outside option can be dependent on the job characteristics, or even endogenously determined by the employment contract. For instance, it is in line with the fact that severance pay is a form of compensation for confidentiality requirements (Yermack 2006). For firms that need to protect their business interests by insisting on executives' confidentiality, it essentially lowers the executives' outside option because terminated executives are unable to fully utilize their human and intellectual capital.

4. Furthermore, our model predicts that the compensation level is increasing with the manager's tenure on the job. In our model, it is because tenure is positively correlated with the agent's past performance (to the extent that empiricists cannot fully control for), and hence with his pay level. This pattern is confirmed by a recent empirical study by Cremers and Palias (2011) (for more discussions on this finding, see the discussion in Section 6.2.2).

5. Last but not least, the salient property of the optimal contract derived in this article is its performance-based back-loaded pattern—that is, the agent gets more cash compensation after a satisfactory performance history. In reality, this core feature is reflected by the performance-based vesting practice in managerial compensation (Bettis et al. 2010), as managers are allowed to cash in these vested stocks/options once they achieve a certain performance target. Another good example is the recent "bonus banking" scheme in which managers can only

[32] Because the agent now has an outside option $\frac{u(z)}{r}$, the admissible continuation payoff W should be above this level. Geometrically, in the state space (W, M) in Figure 3, we need to impose an extra restriction that $W \geq \frac{u(z)}{r}$, which gives our result. Also, the outside option with constant consumption flow z should be interpreted as an unemployment insurance program, and the perpetuity of payment is immaterial. To see this, suppose that the unemployment insurance pays out z only over T years. Then the infinitely lived agent effectively has an outside option of $z' = z\left(1-e^{-rT}\right)$, given his optimal consumption smoothing.

withdraw "banked" bonuses if the firm's subsequent performance remains sound.[33]

6.2.2 Testable empirical predictions. Our theory offers several new empirical predictions on executive compensation, and some of these are (indirectly) supported by existing empirical evidence.

Asymmetric pay-performance sensitivities and related predictions First, our model advocates distinguishing the (cash) pay-performance sensitivities given positive performance shocks from those given negative shocks, and further predicts a positive wedge between the positive cash incentives and the negative ones. In fact, Chen, Liang, and Lin (2006) investigate the asymmetric responses of CEO bonuses following unexpected earnings by running the following regression:

$$\Delta Bonus = \alpha + \beta^{+} \times UE^{+} + \beta^{-} \times UE^{-} + Controls, \qquad (28)$$

where UE^{+} (UE^{-}) is the positive (negative) part of unexpected earnings. Using data from 1993 to 2004, Chen, Liang, and Lin (2006) report a significantly positive bonus incentive wedge $\Delta\beta \equiv \beta^{+} - \beta^{-}$, which says that CEO bonuses increase in a greater magnitude after a positive earning surprise than bonuses decrease after a negative earning surprise. This not only offers support for our model but also shows the empirical relevance of downward-rigid cash compensation. Future empirical research can modify the specification in Equation (28) to be in line with the standard executive compensation literature. For instance, one can replace the unexpected earnings UE by the firm's stock performance, and perhaps incorporate the cash proceeds from exercising options into the explanatory variable.

Our theory suggests the following testable hypothesis. When running the regression in the form of Equation (28), one should expect a greater incentive wedge $\Delta\beta \equiv \beta^{+} - \beta^{-}$ for managers who can easily undo (i.e., smooth out) their on-the-job compensation incentives, because they tend to receive cash compensations that are less sensitive to their downward performance.

The more challenging task is to find proxies for the extent to which managers can smooth out their on-the-job compensation. We can approach this question from two angles. First, by literally interpreting "saving" as consumption smoothing that undoes on-the-job incentives, the manager's age should affect his saving motives due to life-cycle reasons. The life-cycle literature offers some clues in constructing this proxy. For instance, according to Dynan, Edelberg, and Palumbo (2009), the average saving propensity from income

[33] Bonus banking is an incentive plan where part of the bonus earned in a year is "banked" in a bonus account, to be paid out in subsequent years. The firm may declare a negative bonus (sometimes called a "malus") where the amount in the bonus bank is reduced if subsequent corporate or individual performance declines, or if the initial assessment of performance upon which the bonus was based turns out to be wrong.

peaks around age fifty.[34] Therefore, our theory suggests that the incentive wedge $\Delta\beta$ tends to be larger for CEOs with age around fifty.

Second, the flexibility of managers being able to smooth out on-the-job compensation can also be related to firms' corporate governance, which is usually proxied by the index proposed by Gompers, Ishii, and Metrick (2003). Arguably, a company with worse corporate governance will have fewer tools to restrict the managers' undoing activities that blunt their on-the-job cash incentives. If this is true, then our theory predicts that a firm with worse corporate governance should be closer to the bottom panel in Figure 4, and therefore its executive compensation should feature a greater incentive asymmetry $\Delta\beta$.

Corporate governance and pay level Following the interpretation that managers in low-corporate-governance firms have greater flexibility in smoothing out on-the-job compensation, the bottom panel in Figure 4 suggests that these firms should have a stronger positive relationship between pay level and tenure, an empirical pattern documented in Cremers and Palias (2011). Interestingly, this prediction, although perfectly consistent with the entrenchment theory (Bebchuk and Fried 2004), is derived under our model that features optimal contracting with frictions.

Thus, it is nontrivial to tell the entrenchment story apart from the one of optimal contracting with frictions. From a practical point of view, it is likely that poor corporate governance causes a manager to have greater flexibility to undo on-the-job cash incentives. However, caution has to be taken about the origin of friction, as corporate governance is likely to be endogenous.[35] This potential endogeneity issue makes the identification even harder.

Nevertheless, the next prediction may have the power to differentiate our theory from the entrenchment story proposed by Bebchuk and Fried (2004).

Positive association between forced turnovers and options-like contracts The underlying mechanism of our model is as follows. To mitigate the manager's undoing activities through consumption-smoothing, the optimal contract offers downside-protected cash compensation packages. However, to maintain proper incentives, the optimal contract should invoke "sticks" more often, which results in more frequent forced turnovers.

Therefore, our theory suggests that forced turnovers are more likely to occur for managers who receive more options-like compensation packages (which can be measured either directly, or indirectly by the cash incentive wedge $\Delta\beta$

34 This number is available at Figure 2 in their 2008 working paper. Of course, an important caveat is that, compared with their sample, corporate executives may have significantly different saving profiles.

35 Broadly, it is quite possible that both the measured poor corporate governance and the contracting friction (in the model it is the manager's flexibility in undoing his on-the-job cash compensation) are driven by some underlying unobservable factors, and it is these factors that give rise to the increasing pattern of pay level associated with tenure.

as in Equation (28)). Empirically, Jenter and Lewellen (2010) find that boards aggressively fire CEOs for poor performance, and Kaplan and Minton (2008) document that executives' turnover rate has risen since 1998. Because this time period coincides with the one with increasing usage of options in executive compensation, the broad time-series pattern is roughly consistent with our theory. We await future cross-sectional empirical tests on this positive relation between options-like compensation packages and forced turnovers. Perhaps more importantly, this test has the power to differentiate our theory from the entrenchment theory by Bebchuk and Fried (2004), simply because entrenched CEOs should also be less likely to fire themselves following poor performance.

7. Extensions

7.1 Renegotiation-proof contract

In this model, because termination imposes ex post inefficiency, without commitment both parties would like to renegotiate whenever the original contract can be Pareto improved. For the contract to be renegotiation proof, the resulting value function $J^{RP}(W,M)$, where "RP" stands for renegotiation-proof, must be non-increasing in the agent's continuation payoff W. Otherwise, both parties can be strictly better off by raising W.[36]

In Appendix A.10, we construct the value function $J^{RP}(W,M)$ recursively. Analogous to the unidimensional result in DeMarzo and Sannikov (2006), $J^{RP}(W,M)$ features a renegotiation boundary $\underline{W}(M)$ with $J_W\left(\underline{W}(M),M\right) = 0$. The renegotiation curve $\underline{W}(M)$ is the lower bound of the agent's continuation payoff W along the equilibrium employment path at the pay level $c(M)$. When the liquidation value L is relatively high (see Appendix A.10 for detailed conditions), $\underline{W}(M)$ (which might bind at $\frac{U(M)}{r}$) is strictly below the compensation-setting curve $W^*(M)$ (see the left panel in Figure 5).

We have similar results for the renegotiation-proof optimal contract. However, when poor performance drives W down to $\underline{W}(M)$, investors and the agent run a lottery, whose outcome is independent of the cash-flow occurrence. The agent is fired (so W becomes $\frac{U(M)}{r}$ and he loses $\underline{W}(M)-\frac{U(M)}{r}$) with a probability

$$\frac{b+U(M)-r\underline{W}(M)}{\underline{W}(M)-\frac{U(M)}{r}}dt;$$

otherwise, the agent stays at $\underline{W}(M)$. Under this lottery, at $W=\underline{W}(M)$, without success the agent's (expected) dW remains $\left[r\underline{W}(M)-U(M)-b\right]dt$ as in Equation (20). The right panel in Figure 5 gives an example of $J^{RP}(W,M)$

36 The definition of renegotiation-proofness here is the same as in DeMarzo and Fishman (2007), which is equivalent to the contract being sequentially undominated (in terms of both parties' payoffs); see Hart and Tirole (1988). In contrast, Hart and Moore (1998) use a different approach. See related comments in DeMarzo and Fishman (2007).

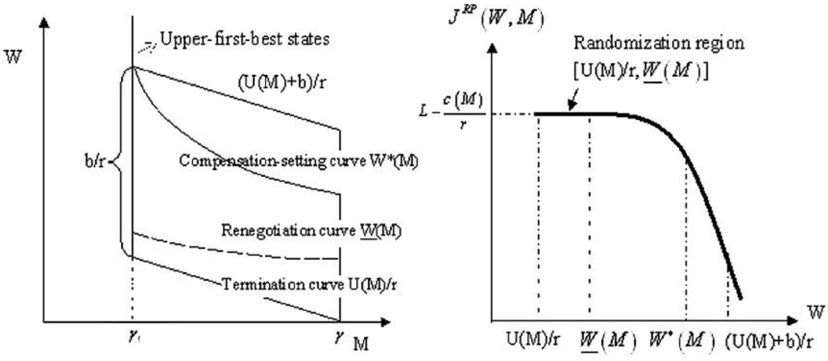

Figure 5
The renegotiation-proof contract
The left panel shows (W, M) space with renegotiation-proof. There exists a renegotiation curve $\underline{W}(M) > \frac{U(M)}{r}$ such that $J_W^{RP}\left(\underline{W}(M), M\right) = 0$, $J_W^{RP}(W, M) < 0$ for $W > \underline{W}(M)$. The right panel shows one example of investors' value function $J^{RP}(W, M)$ as a function of W by fixing M.

as a function of W (fixing M). As shown, $J_W^{RP}(W, M) \leq 0$, and J^{RP} is flat with respect to W in the region of $\left[\frac{U(M)}{r}, \underline{W}(M)\right]$, reflecting the randomization (lottery) between $\frac{U(M)}{r}$ and $\underline{W}(M)$. For detailed constructions and proofs, see Appendix A.10.

7.2 The complete contract with multitasking: A convergence result

We have envisioned the myopic actions as the situation where excessive incentives will lead the agent to hurt the firm with some noncontractible loss Δ. In this sense, our contracting space is *incomplete*. How far away is our optimal contract from the optimal *complete* contract?

To answer this question, we embed a multitasking problem (as in Holmstrom and Milgrom 1991) into the main model. Assume that the firm's operation involves another business activity, which generates a contractible instantaneous *value increment* as

$$dQ_t = -\Delta \mathbf{1}_{\{a_t = \overline{p}\}} dt + \sigma dZ_t,$$

where $\{Z_t\}$ is a Brownian process independent of $\{N_t\}$. We may also interpret dQ_t as the (noisy) change of the firm's long-run value. Neither shirking nor working has any impact on the drift in dQ_t. Once the agent takes the myopic action $a = \overline{p}$, however, the drift becomes $-\Delta$ as the agent transfers his effort allocation from the soft performance dQ_t to the hard performance dN_t. Due to the risk-neutrality of investors, if the resulting complete contract does ignore dQ_t completely, then we are back to the contracting space considered in Section 4.

When the loss Δ is contractible through dQ_t, investors can raise the incentive loading on dN_t but still prevent the agent from taking myopic actions.

The contract can specify an incentive scheme such as

$$dW_t = (rW_{t^-} - U(M_{t^-}))\,dt + \beta_t^W (dN_t - pdt) + x_t dQ_t, \qquad (29)$$

where the incentive loading on the hard performance dN_t is $\beta_t^W = \frac{b}{p} + k_t > \frac{b}{p}$ and $k_t > 0$, and the incentive loading on the soft performance dQ_t is x_t. Now, if we set

$$x_t = \frac{k_t(\overline{p} - p)}{\Delta} = \frac{k_t \epsilon}{\Delta},$$

then the agent will be refrained from the myopic action: by taking $a = \overline{p} = p + \epsilon$, the agent gains $k_t \epsilon$ from dN_t, but this gain is offset by the loss $x_t \Delta$ from dQ_t.

As discussed in Section A.8.2, setting $\beta_t^W > \frac{b}{p}$ (or equivalently $k_t > 0$) gives rise to a benefit in relaxing the no-savings constraint, as investors may cut compensation on the path without any success. However, it is costly to set $k_t > 0$, and in turn $x_t = \frac{k_t \epsilon}{\Delta} > 0$. This is because by imposing positive loading $x_t > 0$ on the agent's continuation payoff in Equation (29), the noise in dQ_t makes inefficient terminations more likely. In addition, it is also inefficient to expose the risk-averse agent to random noises. Based on this observation, He (2008) shows that when the information precision of dQ_t goes to zero (i.e., $\sigma \to \infty$ to capture the softness of dQ as in Holmstrom and Milgrom 1991), the value from the optimal complete contract converges to the one from the incomplete contract derived in Section 4, and investors tend to ignore such extremely noisy signals (i.e., $x_t \to 0$). This theoretical result implies that the "incomplete" contract derived in Section 4 can be optimal even in the paradigm of complete contracts, if the information dQ is sufficiently "soft" and there exists some positive transaction cost in procuring this soft information.

7.3 General utility functions

The adoption of CARA utility is only for exposition purposes. This section extends our analysis to a general utility function $u(\cdot)$ that satisfies condition (2). Similar to (17), by writing $g(c) = u'(c)$, we define the agent's utility as a function of the marginal utility M to be $U(M) = u\left(g^{-1}(M)\right)$. Now the termination boundary $l(M) = \frac{U(M)}{r}$ is no longer a line as in the CARA case (see Figure 3). For the concavity of the value function J, we require the domain $\{(W, M) : W \geq l(M)\}$ to remain convex, which implies that $l(M)$ is a convex function. One can easily check that $l(M)$ is convex if and only if $u''' > \frac{(u'')^2}{u'}$, a property that is also satisfied by the class of power utility.

The structure of resulting optimal contract remains unchanged: the compensation process $\{c\}$ is nondecreasing before M reaches its lower bound; the agent works for potential pay raises; and the agent's poor performance leads to dismissal, but he walks away with a severance payment $\frac{c_\tau}{r}$. Interested readers can find detailed constructions for the general utility case in He (2008).

8. Conclusion

We study a dynamic agency problem in which the agent can save privately. When ruling out private savings, previous studies (Rogerson 1985; Sannikov 2008; etc.) derive a front-loaded, performance-sensitive compensation flow in the optimal contract. In contrast, the optimal compensation process in this article is back-loaded and relatively insensitive to poor performance, and the agent may be dismissed with a severance package. Our theory can simultaneously explain the popularity of options-like compensation contracts and the increasing incidence of forced turnovers with sizeable severance pay.

We solve the optimal contracting problem with private savings by utilizing the binding incentive-compatibility constraint in the presence of myopic actions, where the linearity of effort cost structure is important. However, in justifying the noncontractibility via information acquisition costs in Section 7.2, we employ a proof method that allows the agent's cost structure to be convex, and we show the convergence result when the convexity diminishes (for details, see He 2008). Therefore, our contracting result is a general one in this regard.

We emphasize that the resulting contract form, especially the perfect downward-rigidity, is specific to our particular setting. As suggested in Remark 3, a less-responsive compensation pattern and a positive severance pay, which are designed to reduce the agent's deviation values, should be robust features of the optimal contract when the agent can privately save. The exact degree of robustness needs future theoretical work to explore more general settings, which might give further guidelines in solving the optimal contracting problems with private savings.

A. Appendix

A.1 Proof of Lemma 1

Suppose that under $\Pi = \{\{c\}, F_\tau, \tau\}$, the agent's optimal consumption-saving strategy is $\{\widehat{c}_t \neq c_t; S_t \geq 0\}$. Consider offering the contract $\widehat{\Pi} = \left\{\{\widehat{c}\}, \widehat{F}_\tau = \frac{\widehat{c}_\tau}{r}, \tau\right\}$ (instead of Π) to the agent; clearly, this contract just replicates the agent's optimal consumption profile under Π. Now we show that the agent will not deviate under the new contract $\widehat{\Pi}$. Suppose not; then, there is a saving path $\left\{\widehat{S} \geq 0\right\}$ combined with another action profile $\left\{a'\right\}$ to support a consumption profile $\left\{c'\right\}$ that achieves a strictly higher value for the agent. But then $\left\{S' = S + \widehat{S} \geq 0\right\}$ with the action profile $\left\{a'\right\}$ can support $\left\{c'\right\}$ under the original contract $\Pi = \{\{c\}, F_\tau, \tau\}$, in contradiction to the optimality of $\{\widehat{c}_t \neq c_t; S_t \geq 0\}$ under the original contract Π.

A.2 Proof of Proposition 1

Take the zero-saving policy as given. Under the preassumption that $a_t = p$ for all t, the agent's value process is

$$V_t = \mathbb{E}_t\left[\int_0^\tau e^{-rt} u(c_t)\, dt + e^{-r\tau}\frac{u(c_\tau)}{r}\right],$$

and the martingale representation theorem (e.g., Biais et al. 2007) implies that there exists an $\mathcal{F}^N$ predictable process $\left\{\beta_s^W\right\}$ such that

$$V_t = V_0 + \int_0^t e^{-rs}\beta_s^W\left(-pds + dN_s\right).$$

According to the definition of W_t, we have

$$V_t = \int_0^t e^{-rs}u\left(c_s\right)ds + e^{-rt}W_t;$$

and then, differentiating both sides, we obtain the expression in Equation (7).

Now consider any feasible effort process $a = \{a_t \in \{0, p, \overline{p}\} : t \in [0, \tau)\}$. The agent's associated value process $V_t\left(a\right)$ could be written as

$$V_t\left(a\right) = V_0 + \int_0^t e^{-rs}\beta_s^W\left(-pds + dN_s\left(a_s\right)\right) + \int_0^t e^{-rt}\frac{b}{p}\left(p - a_s\right)ds,$$

where $dN_s\left(a_s\right)$ has an intensity of a_s. Then,

$$\begin{aligned} dV_t\left(a\right) &= e^{-rt}\beta_s^W\left(-pdt + dN_t\left(a_t\right)\right) - e^{-rt}\frac{b}{p}\left(a_t - p\right)dt \\ &= e^{-rt}\left(a_t - p\right)\left(\beta_s^W - \frac{b}{p}\right)dt + e^{-rt}\beta_s^W\left(dN_t\left(a_t\right) - a_t dt\right). \end{aligned}$$

Therefore, to implement working, it must be the case that $\left(a_t - p\right)\left(\beta_s^W - \frac{b}{p}\right) \leq 0$ for both $a_t = 0$ and $a_t = \overline{p}$. This implies that $\beta_s^W = \frac{b}{p}$, a binding incentive-compatibility constraint. It directly follows that the agent obtains the same value by taking any action process $\{a\}$ s.t. $a_t = 0$ or p. Q.E.D.

A.3 Proof of Proposition 2

The first result is just Proposition 1. Now we prove the second result. Note that in the following proof we allow for randomization other than the agent's Poisson performance in the contract. Suppose not—then the contract must specify some paths on $[0, \tau)$ with strictly positive measure so that $\mathbb{E}_t^a\left[M_{t'}\right] > M_t$ for some action process a and $t' > t$. Collect these time points into a set T with positive Lebesgue measure (in time), so that on this set T (indexed by the element $t \in T$) the marginal utility follows a submartingale (in expectation it is increasing).

Now consider the following profitable consumption-smoothing strategy on this set T, in which the agent saves a bit in the beginning of T and consumes in the end of T. Pick the lowest (highest) t's to form T_l $(T_h) \subset T$ so that the Lebesgue measure of T_l is $\varepsilon > 0$, where T_l (T_h) has a higher (lower) marginal utility. Choose ε to be sufficiently small so that

$$t_l^1 \equiv \inf T_l < t_l^2 \equiv \sup T_l < t_h^1 \equiv \inf T_h < t_h^2 \equiv \sup T_h,$$

and without loss of generality we set $t_l^1 = \inf T_l = \inf T = 0$. At t_l^2, the agent's marginal utility (wage) is strictly lower (higher) than that at t_h^1 so that

$$\mathbb{E}_{t=0}^a\left(M_{t_l^2}\right) < \mathbb{E}_{t=0}^a\left(M_{t_h^1}\right). \tag{A1}$$

Otherwise, $\mathbb{E}^a_{t=0}\left(M_{t_l^2}\right) = \mathbb{E}^a_{t=0}\left(M_{t_h^1}\right)$ for any small $\varepsilon > 0$, plus the fact that M follows a submartingale, immediately imply that M_t is martingale on set T a.e., contradiction to the construction of T.

Suppose now that the agent saves $e^{rt}\epsilon$ for $t \in T_l$ and consumes $e^{rt'}\epsilon$ for $t' \in T_h$; clearly, this satisfies the savings technology (if at T_l in some states wages are zero, then only consider saving on the states with strictly positive wages, and consume these savings at T_h). The total utility loss from lowering consumption on T_l is

$$\mathbb{E}^a_{t=0}\left[\int_{T_l} e^{-rt}\left(u\left(c_t\right) - u\left(c_t - e^{rt}\epsilon\right)\right)dt\right] = \mathbb{E}^a_{t=0}\int_{T_l}\left[M_t\epsilon + o\left(\epsilon\right)\right]dt < \epsilon\varepsilon\mathbb{E}^a_{t=0}\left(M_{t_l^2}\right) + o\left(\epsilon\right),$$

because $\mathbb{E}^a_{t=0}\left(M_t\right) < \mathbb{E}^a_{t=0}\left(M_{t_l^2}\right)$ on $t \in T_l$. Similarly, the utility gain from raising consumption on T_h is

$$\begin{aligned}\mathbb{E}^a_{t=0}\left[\int_{T_h} e^{-rt}\left(u\left(c_t + e^{rt}\epsilon\right) - u\left(c_t\right)\right)dt\right] &= \mathbb{E}^a_{t=0}\int_{T_h}\left[M_t\epsilon + o\left(\epsilon\right)\right]dt \\ &> \epsilon\varepsilon\mathbb{E}^a_{t=0}\left(M_{t_h^1}\right) + o\left(\epsilon\right).\end{aligned}$$

Therefore, the total gain

$$\begin{aligned}&\mathbb{E}^a_{t=0}\left[\int_{T_h} e^{-rt}\left(u\left(c_t + e^{rt}\epsilon\right) - u\left(c_t\right)\right)dt\right] - \mathbb{E}^a_{t=0}\left[\int_{T_l} e^{-rt}\left(u\left(c_t\right) - u\left(c_t - e^{rt}\epsilon\right)\right)dt\right] \\ &= \epsilon\varepsilon\left[\mathbb{E}^a_{t=0}\left(M_{t_h^1}\right) - \mathbb{E}^a_{t=0}\left(M_{t_l^2}\right)\right] + o\left(\epsilon\right).\end{aligned}$$

When ϵ is sufficiently small, this is dominated by the first term, which is strictly positive due to Equation (A1).

A.4 Proof of Lemma 2

The "if" part is obvious. Now we prove the "only if" part. Let us take the equilibrium effort process $\{a = p\}$ first. Then, according to the Doob-Meyer decomposition theorem (see, e.g., Karatzas and Shreve 1988) and the martingale representation theorem (see, e.g., Biais et al. 2007), there exist an $\mathcal{F}^N$-predictable process $\left\{\beta_t^M\right\}$ and a predictable non-increasing process $\left\{H_t^D\right\}$ such that

$$dM_t = dH_t + \beta_t^M\left(dN_t - pdt\right),$$

with $dH_t \le 0$. Define $dM_t^D \equiv dH_t - \beta_t^M pdt$ so that M_t^D is also a predictable process. Then

$$dM_t = dM_t^D + \beta_t^M dN_t,$$

and since $dH_t \le 0$ we have $dM_t^D \le -\beta_t^M pdt$. We need to further prove that $dM_t^D \le 0$. Suppose it is not; say that $dM_t^D > 0$ holds for some paths with strictly positive measure. Then, if the agent takes the effort $a = 0$ on these paths, $dN_t = 0$, and dM_t is strictly increasing on these paths with strictly positive measure. This contradicts with condition 2 in Proposition 2.

A.5 Proof of Lemma 3

Clearly, firing the agent delivers the continuation payoff of $W=\frac{U(M)}{r}$. Now we show that there are no other ways to deliver $W=\frac{U(M)}{r}$. We have two steps to go, and in the following argument t can be understood as t^- as the information at $(t-dt,t]$ is irrelevant.

(1) Note that to respect condition (9), given a marginal utility M, any continuation payoff $W<\frac{U(M)}{r}$ is infeasible. The argument is as follows. In light of Proposition 2, for any equilibrium effort policy a, no savings implies that $M_t \geq \mathbb{E}_t^a(M_s)$ for $s>t$ (s could be larger than τ, in which case the agent is fired and the distribution is degenerate). According to the definition of W_t, which is the agent's optimal value, we have

$$W_t \geq \mathbb{E}_t^a\left[\int_t^{\infty} e^{-r(s-t)}U(M_s)\,dt\right] \geq \int_t^{\infty} e^{-r(s-t)}U\left(\mathbb{E}_t^a M_s\right)dt \geq \frac{U(M_t)}{r},$$

where the first "≥" is due to the possibility of $M_s=\gamma_L$, the second "≥" is due to the convexity of $U(\cdot)$ (in the CARA case it is a linear function; see related discussions in Section 7.3), and the third "≥" is because $U(\cdot)$ is decreasing.

(2) The necessary condition (13) implies that at the point $W=\frac{U(M)}{r}$, W is a martingale. Since W cannot fall, it has to remain constant $\frac{U(M)}{r}$ from then on. Because the agent obtains the same payoff by shirking and working, this implies zero potential shirking benefit. Therefore, in this case, the agent is fired.

A.6 Appendix for Section 4

Given K, the total number of potential cash flows, we use $i\leq K$ to indicate the number of cash flows remaining, and we are solving for $J^{i,K}$. But in our setting, since only the number of cash flows remaining matters, $J^{i,K}$ is independent of K. Therefore, we omit K in the following analysis.

A.6.1 Production Stage When $i=0$, there are no future cash flows, and the firm is obsolete. Based on the definition of $J^L(W)$ in Equation (19), we have

$$J^0(W,M)=\begin{cases} J^L(W) \text{ if } W\geq \frac{U(M)}{r} \\ -\infty \quad\quad \text{otherwise}\end{cases}.$$

It is clear that $J^0(W,M)$ satisfies all conditions in Proposition 4. Now consider $i\geq 1$. The next lemma translates Proposition 4 to the corresponding properties of j^{i-1}.[37]

Lemma 4. For the compensation-setting stage value function j^{i-1}, we have the following properties:

1. $j_w^{i-1}\geq -\frac{1}{\gamma_L}$, and $\frac{1}{\gamma rm}<j_m^{i-1}\leq\frac{1}{\gamma r\gamma_L}$.

[37] Strictly speaking, here, all the second-order derivatives—j_{ww}, j_{wm}, and j_{mm}—are in the weak sense (in a Soboslov space), which allows for (finite) discontinuities, and the integration-by-parts formula still holds. To be precise, in the production stage $\tilde{j}^i$ is a mollified version of j^{i-1}, which makes everything smooth, but the compensation-setting stage only keeps the first-order smoothness (that is, for the second-order derivatives there will be a discontinuity on $M^*(W)$). However, because the first-order derivatives are continuous, the negative definiteness of Hessian matrix is sufficient for the concavity.

2. $j_{ww}^{i-1} < 0$, $j_{mm}^{i-1} < 0$, $j_{wm}^{i-1} > 0$ and $j_{ww}^{i-1} j_{mm}^{i-1} - \left(j_{wm}^{i-1}\right)^2 \geq 0$. Therefore, $j^{i-1}(w, m)$ is concave.
3. $\frac{1}{\gamma r} j_{ww}^{i-1} + j_{wm}^{i-1} < 0$, $\frac{1}{\gamma r} j_{w}^{i-1} + j_{m}^{i-1} \geq 0$, and $\frac{1}{\gamma r} j_{w}^{i-1}\left(\frac{b}{r}, m\right) + j_{m}^{i-1}\left(\frac{b}{r}, m\right) = 0$.

We carry out our analysis based on the following linear transformation:

$$\begin{cases} w = W - \frac{U(M)}{r} \in \left[0, \frac{b}{r}\right], \\ m = M \in [\gamma_L, \gamma], \end{cases}$$

where the domain is a rectangle. Let $\widetilde{j}^i(w, m) = \widetilde{J}^i(W, M)$, and $j^i(w, m) = J^i(W, M)$. Clearly, $\widetilde{j}\,(j)$ is concave if and only if $\widetilde{J}\,(J)$ is concave. Note that

$$\widetilde{J}_W^i = \widetilde{j}_w^i,\ \widetilde{J}_M^i = \frac{1}{\gamma r}\widetilde{j}_w^i + \widetilde{j}_m^i, \text{ and } \widetilde{J}_{WM}^i = \frac{1}{\gamma r}\widetilde{j}_{ww}^i + \widetilde{j}_{wm}^i,$$

and similar relations hold between j and J.

Without jump, $\widetilde{j}^i$ satisfies the following ODE:

$$(r + p)\,\widetilde{j}^i(w, m) = -c(m) + p\left(Y + j^{i-1}\left(w + \frac{b}{p}, m\right)\right) + j_w(w, m)(rw - b), \quad \text{(A2)}$$

and its closed-form solution is

$$\widetilde{j}^i(w, m) = \frac{r}{r + p} J^L\left(\frac{U(m)}{r}\right) + \frac{p}{r + p}[b - rw]^{1 + \frac{p}{r}} \left[\int_0^w \frac{(r + p)\left(Y + j^{i-1}\left(x + \frac{b}{p}, m\right)\right)}{[b - rx]^{2 + \frac{p}{r}}} dx + \frac{J^L\left(\frac{U(m)}{r}\right)}{b^{1 + \frac{p}{r}}}\right], \quad \text{(A3)}$$

where we use $c(m) = -rJ^L\left(\frac{U(m)}{r}\right)$. The solution in Equation (23) in the main text is identical to Equation (A3).

Based on lemma 4, we have the following lemma for $\widetilde{j}^i$, and the results regarding $\widetilde{J}$ in Lemma 3 follow directly from this lemma.

Lemma 5. For the production-stage value function $\widetilde{j}^i$, we have the following properties:

1. $\widetilde{j}_w^i \geq -\frac{1}{\gamma_L}$, and $\frac{1}{\gamma r m} < \widetilde{j}_m^i \leq \frac{1}{\gamma r \gamma_L}$;
2. $\widetilde{j}_{ww}^i < 0$, $\widetilde{j}_{mm}^i < 0$, $\widetilde{j}_{wm}^i > 0$, and $\widetilde{j}_{ww}^i \widetilde{j}_{mm}^i - \left(\widetilde{j}_{wm}^i\right)^2 > 0$;
3. $\frac{1}{\gamma r}\widetilde{j}_{ww}^i + \widetilde{j}_{wm}^i < 0$, and $\frac{1}{\gamma r}\widetilde{j}_w^i\left(\frac{b}{r}, m\right) + \widetilde{j}_m^i\left(\frac{b}{r}, m\right) < 0$.

Proof. From Equation (A3), it is easy to calculate (note that $\frac{dJ^L\left(\frac{U(m)}{r}\right)}{dm} = \frac{1}{\gamma r m}$)

$$\widetilde{j}_m^i = \frac{r}{r + p}\frac{1}{\gamma r m} + \frac{p}{r + p}[b - rw]^{1 + \frac{p}{r}} \left[\int_0^w (r + p)\, j_m^{i-1}\left(x + \frac{b}{p}, m\right) \times [b - rx]^{-2 - \frac{p}{r}}\, dx + \frac{1}{\gamma r m} b^{-1 - \frac{p}{r}}\right].$$

Notice that

$$\frac{r}{r+p}+\frac{p}{r+p}\left[b-rw\right]^{1+\frac{p}{r}}\left[\int_0^w (r+p)\left[b-rx\right]^{-2-\frac{p}{r}}dx+b^{-1-\frac{p}{r}}\right]=1,$$

which constitutes a probability measure. Since $j_m^{i-1}\in\left[\frac{1}{\gamma rm},\frac{1}{\gamma r\gamma_L}\right]$, we have $\widetilde{j}_m^i\in\left[\frac{1}{\gamma rm},\frac{1}{\gamma r\gamma_L}\right]$.

Based on Equations (A2) and (A3), a direct calculation (where we use the integration-by-parts formula) yields

$$\begin{aligned}&\widetilde{j}_w^i-\frac{p}{b-rw}\left(Y+j^{i-1}\left(w+\frac{b}{p},m\right)\right)\\&=-p\left[b-rw\right]^{\frac{p}{r}}\left[\int_0^w (r+p)\left(Y+j^{i-1}\left(x+\frac{b}{p},m\right)\right)(b-rx)^{-2-\frac{p}{r}}dx+J^L\left(\frac{U(m)}{r}\right)b^{-1-\frac{p}{r}}\right]\\&=p\left[b-rw\right]^{\frac{p}{r}}\left\{\begin{array}{c}\left[Y+j^{i-1}\left(x+\frac{b}{p},m\right)\right](b-rx)^{-1-\frac{p}{r}}\Big|_w^0\\+\int_0^w j_w^{i-1}\left(x+\frac{p}{b}\right)\left[b-rx\right]^{-1-\frac{p}{r}}dx-J^L\left(\frac{U(m)}{r}\right)b^{-1-\frac{p}{r}}\end{array}\right\};\end{aligned}$$

therefore,

$$\begin{aligned}\widetilde{j}_w^i&=p\left[b-rw\right]^{\frac{p}{r}}\left\{\int_0^w j_w^{i-1}\left[b-rx\right]^{-1-\frac{p}{r}}dx\right.\\&\quad\left.+\left[Y+j^{i-1}\left(\frac{b}{p},m\right)-J^L\left(\frac{U(m)}{r}\right)\right]b^{-1-\frac{p}{r}}\right\} &&\text{(A4)}\\&\quad>\left[b-rw\right]^{\frac{p}{r}}\left\{\int_0^w pj_w^{i-1}\left[b-rx\right]^{-1-\frac{p}{r}}dx+j_w^{i-1}\left(\frac{b}{p},m\right)b^{-\frac{p}{r}}+pYb^{-1-\frac{p}{r}}\right\}. &&\text{(A5)}\end{aligned}$$

The second inequality follows from the following fact: $J^L\left(\frac{U(m)}{r}\right)=j^{i-1}(0,m)$, and j^{i-1} is concave, which implies that $j^{i-1}\left(\frac{b}{p},m\right)-J^L\left(\frac{U(m)}{r}\right)>j_w^{i-1}\left(\frac{b}{p},m\right)\cdot\frac{b}{p}$. Since

$$\left[b-rw\right]^{\frac{p}{r}}\left[\int_0^w p\left[b-rx\right]^{-1-\frac{p}{r}}dx+b^{-\frac{p}{r}}\right]=1, \tag{A6}$$

which constitutes a probability measure, from Equation (A5) we know that $\widetilde{j}_w^i>j_w^{i-1}\geq-\frac{1}{\gamma_L}$. Also, in the limiting case $w=\frac{b}{r}$, we have

$$\widetilde{j}_w^i\left(\frac{b}{r},m\right)=j_w^{i-1}\left(\frac{b}{r}+\frac{b}{p},m\right), \tag{A7}$$

simply because when $w\to\frac{b}{r}$, the entire probability weights in Equation (A6) are put on $w=\frac{b}{r}$.

Now we study the second-order derivatives. It is straightforward that

$$\begin{aligned}-\widetilde{j}_{mm}^i&=\frac{r}{r+p}\frac{1}{\gamma rm^2}+\frac{p}{r+p}\left[b-rw\right]^{1+\frac{p}{r}}\left[\int_0^w\frac{(r+p)\left[-j_{mm}^{i-1}\left(x+\frac{b}{p},m\right)\right]}{\left[b-rx\right]^{2+\frac{p}{r}}}dx+\frac{1}{\gamma rm^2}b^{-1-\frac{p}{r}}\right]\\&=\left[b-rw\right]^{1+\frac{p}{r}}\left[\int_0^w\frac{p\left[-j_{mm}^{i-1}\left(x+\frac{b}{p},m\right)\right]}{\left[b-rx\right]^{2+\frac{p}{r}}}dx+\left[\frac{p}{r+p}+\frac{r}{r+p}\left(\frac{b}{b-rw}\right)^{1+\frac{p}{r}}\right]\frac{b^{-1-\frac{p}{r}}}{\gamma rm^2}\right]>0.\end{aligned}$$

This shows that $\widetilde{j}^i$ is concave in m. For $\widetilde{j}^i_{ww}$, we use Equations (A2) and (A5), and find that

$$-\widetilde{j}^i_{ww} = \frac{p}{b-rw}\left(\widetilde{j}^i_w - j^{i-1}_w\left(w+\frac{b}{p},m\right)\right) > [b-rw]^{\frac{p}{r}-1} b^{-1-\frac{p}{r}} p^2 Y$$
$$+\frac{p}{b-rw}\left[[b-rw]^{\frac{p}{r}}\left[\int_0^w p j^{i-1}_w [b-rx]^{-1-\frac{p}{r}}\,dx + j^{i-1}_w\left(\frac{b}{p},m\right) b^{-\frac{p}{r}}\right]\right.$$
$$\left. - j^{i-1}_w\left(w+\frac{b}{p},m\right)\right].$$

Invoking the integration-by-parts technique again, we have

$$[b-rw]^{\frac{p}{r}}\left[\int_0^w p j^{i-1}_w [b-rx]^{-1-\frac{p}{r}}\,dx + j^{i-1}_w\left(\frac{b}{p},m\right) b^{-\frac{p}{r}}\right]$$
$$= j^{i-1}_w\left(w+\frac{b}{p},m\right) + [b-rw]^{\frac{p}{r}}\int_0^w \left(-j^{i-1}_{ww}\right)[b-rx]^{-\frac{p}{r}}\,dx,$$

and therefore

$$-\widetilde{j}^i_{ww} > [b-rw]^{\frac{p}{r}-1} b^{-1-\frac{p}{r}} p^2 Y + p\,[b-rw]^{\frac{p}{r}-1}\int_0^w \left(-j^{i-1}_{ww}\right)[b-rx]^{-\frac{p}{r}}\,dx > 0. \tag{A8}$$

Shortly we will need a stronger estimate for the global concavity of $\widetilde{j}$. According to condition (24), $Y > \frac{1}{\gamma r}\left[\frac{\gamma}{\gamma_L}-1\right]^2$, and

$$-\widetilde{j}^i_{ww} > [b-rw]^{\frac{p}{r}-1}\left[\int_0^w p\left(-j^{i-1}_{ww}\right)[b-rx]^{-\frac{p}{r}}\,dx + \frac{p^2}{b^2\gamma r}\left[\frac{\gamma}{\gamma_L}-1\right]^2 b^{-\frac{p}{r}+1}\right].$$

Finally, we calculate

$$\widetilde{j}^i_{wm} = \frac{\partial}{\partial m}\widetilde{j}^i_w = [b-rw]^{\frac{p}{r}}\left[\int_0^w p j^{i-1}_{wm}[b-rx]^{-1-\frac{p}{r}}\,dx\right.$$
$$\left. +\frac{p}{b}\left[j^{i-1}_m\left(\frac{b}{p},m\right) - \frac{1}{\gamma r m}\right] b^{-\frac{p}{r}}\right]; \tag{A9}$$

it immediately implies that $\widetilde{j}^i_{wm} \geq 0$, because $j^{i-1}_{wm} \geq 0$ and $j^{i-1}_m\left(\frac{b}{p},m\right) > j^{i-1}_m(0,m) = \frac{1}{\gamma r m}$.

Now we show that $\widetilde{j}^i$, in fact, is globally concave, which requires that $\widetilde{j}^i_{ww}\widetilde{j}^i_{mm} > \left(\widetilde{j}^i_{wm}\right)^2$. To show this, we invoke the Cauchy-Schwartz inequality. Observe that the terms other than the integral in $\widetilde{j}^i_{ww}$, $\widetilde{j}^i_{mm}$, and $\widetilde{j}^i_{wm}$ are $\frac{p^2}{b^2\gamma r}\left[\frac{\gamma}{\underline{\gamma}}-1\right]^2$, $\left[\frac{p}{r+p}+\frac{r}{r+p}\left(\frac{b}{b-rw}\right)^{1+\frac{p}{r}}\right]\frac{1}{\gamma r m^2} \geq \frac{1}{\gamma r m^2}$, and $\frac{p}{b}\left[j^{i-1}_m\left(\frac{b}{p},m\right) - \frac{1}{\gamma r m}\right]$, respectively, and we have

$$\frac{p^2}{b^2\gamma r}\left[\frac{\gamma}{\gamma_L}-1\right]^2\left[\frac{p}{r+p}+\frac{r}{r+p}\left(\frac{b}{b-rw}\right)^{1+\frac{p}{r}}\right]\frac{1}{\gamma r m^2}$$
$$> \frac{p^2}{b^2\gamma^2 r^2}\left[\frac{1}{\gamma_L}-\frac{1}{m}\right]^2 > \frac{p^2}{b^2}\left[j^{i-1}_m\left(\frac{b}{p},m\right)-\frac{1}{\gamma r m}\right]^2.$$

Then, the standard Cauchy-Schwartz argument yields that

$$\widetilde{j}^i_{ww}\widetilde{j}^i_{mm} > [b-rw]^{\frac{2p}{r}}\left[\int_0^w p\left(j^{i-1}_{ww}j^{i-1}_{mm}\right)^{\frac{1}{2}}[b-rx]^{-1-\frac{p}{r}}dx+\frac{p}{b}\left[j^{i-1}_m\left(\frac{b}{p},m\right)-\frac{1}{\gamma rm}\right]b^{-\frac{p}{r}}\right]^2$$
$$> [b-rw]^{\frac{2p}{r}}\left[\int_0^w p\left|j^{i-1}_{wm}\right|[b-rx]^{-1-\frac{p}{r}}dx+\frac{p}{b}\left[j^{i-1}_m\left(\frac{b}{p},m\right)-\frac{1}{\gamma rm}\right]b^{-\frac{p}{r}}\right]^2 \geq \left(\widetilde{j}^i_{wm}\right)^2,$$

where we use the fact that j^{i-1} is concave.

Finally, we show property 3. According to condition (24), $Y > \frac{b}{p\gamma_L}$. Utilizing Equations (A9) and (A8), and since $j^{i-1}_m\left(\frac{b}{p},m\right)-\frac{1}{\gamma rm} < \frac{1}{\gamma r}\frac{1}{\gamma_L}$, we have

$$\begin{aligned}\frac{1}{\gamma r}\widetilde{j}^i_{ww}+\widetilde{j}^i_{wm} &< [b-rw]^{\frac{p}{r}-1}\left[\int_0^w p\left(\frac{1}{\gamma r}j^{i-1}_{ww}\right)[b-rx]^{-\frac{p}{r}}dx\right]\\ &\quad+[b-rw]^{\frac{p}{r}}\left[\int_0^w pj^{i-1}_{wm}[b-rx]^{-1-\frac{p}{r}}dx\right]\\ &= [b-rw]^{\frac{p}{r}}\left[\int_0^w p\left(\frac{1}{\gamma r}j^{i-1}_{ww}+j^{i-1}_{wm}\right)[b-rx]^{-1-\frac{p}{r}}dx\right]\\ &\quad+[b-rw]^{\frac{p}{r}-1}\left[\int_0^w p\left(\frac{1}{\gamma r}j^{i-1}_{ww}\right)[b-rx]^{-\frac{p}{r}}dx\right]\\ &\quad-[b-rw]^{\frac{p}{r}}\left[\int_0^w p\left(\frac{1}{\gamma r}j^{i-1}_{ww}\right)[b-rx]^{-1-\frac{p}{r}}dx\right].\end{aligned}$$

The first item is negative because $\frac{1}{\gamma r}j^{i-1}_{ww}+j^{i-1}_{wm} < 0$. Because $j^{i-1}_{ww} < 0$, and for $x < w$ we have

$$[b-rw]^{\frac{p}{r}-1}[b-rx]^{-\frac{p}{r}} > [b-rw]^{\frac{p}{r}}[b-rx]^{-1-\frac{p}{r}},$$

the second item is negative too. Therefore, $\frac{1}{\gamma r}\widetilde{j}^i_{ww}+\widetilde{j}^i_{wm} < 0$.

The second inequality in property 3 says that $\widetilde{J}^i_M\left(\frac{U(M)+b}{r}+\frac{b}{p},M\right) < 0$. To show this, when $W=\frac{U(M)+b}{r}$, we take the derivative with respect to M on Equation (23) to obtain that

$$r\widetilde{J}^i_M = -c'(M)+p\left(J^{i-1}_M\left(\frac{U(M)+b}{r}+\frac{b}{p},M\right)-\widetilde{J}^i_M\right)+\frac{1}{\gamma}\widetilde{J}^i_W \Rightarrow \widetilde{J}^i_M$$
$$=\frac{1}{\gamma(r+p)}\left(\frac{1}{M}+\widetilde{J}^i_W\right),$$

where we use $J^{i-1}_M\left(\frac{U(M)+b}{r}+\frac{b}{p},M\right)=0$ (Proposition 4, property 3), and $-c'(M)=\frac{1}{\gamma M}$. However, we have shown that $\widetilde{J}^i_W\left(\frac{U(M)}{r}+\frac{b}{r},M\right)=j^{i-1}_w\left(\frac{b}{r}+\frac{b}{p},m\right)$ in Equation (A7). Now, as verified in the next compensation-setting stage, investors raise the agent's wage, and as a result there exists $m^* < m$ so that

$$j^{i-1}_w\left(\frac{b}{r}+\frac{b}{p},m\right) = -\gamma rj^{i-1}_m\left(\frac{b}{r}+\frac{U(m)-U\left(m^*\right)}{r}+\frac{b}{p},m^*\right)$$
$$< -\gamma r\frac{1}{\gamma rm^*} \leq -\frac{1}{m}.$$

Therefore, $\widetilde{J}^i_W < -\frac{1}{M}$, and $\widetilde{J}^i_M\left(\frac{U(M)+b}{r},M\right) < 0$. Q.E.D. ■

A.6.2 Compensation-setting stage First we show the zero marginal cost brought on by the future termination at $W = \frac{b+U(M)}{r}$. Notice that raising that wage at $W = \frac{b+U(M)}{r}$ is equivalent to setting w below $\frac{b}{r}$. Consider the policy of setting $W^*(M)$ so that $w = \frac{b}{r} - \varepsilon$. Then, starting from $\left(W^*(M), M\right)$, it is easy to check that the expected discounted termination probability is $\left(\frac{r\varepsilon}{b}\right)^{\frac{r+p}{r}}$ on the path without any jumps. Under the Poisson setup, the total expected discounted termination probability—by integrating over all jumps but with the same M—is still in the order of $\varepsilon^{\frac{r+p}{r}}$; notice that it is an upper-bound estimator, as if a jump leads to a lower M^*, then the impact on the probability of future terminations is zero. Therefore, reducing M has a zero marginal impact on $\varepsilon = 0$ when $p > 0$.

Next we present a formal construction of J^i from $\widetilde{J}^i$. Given $M^*(W)$ defined in the main text (note that $M^*(\cdot)$ might be i-dependent), we propose a transformation

$$\mathbb{T}(W, M) = \left(W, \min\left(M, M^*(W)\right)\right), \tag{A10}$$

and define $J^i(W, M) = \widetilde{J}^i(\mathbb{T}(W, M))$. This transformation preserves the concavity. To see this, consider any two points (W_1, M_1) and (W_2, M_2) and

$$W(\lambda) = \lambda W_1 + (1-\lambda) W_2 \text{ and } M(\lambda) = \lambda M_1 + (1-\lambda) M_2.$$

For $\mathbf{S} = \mathbb{T}(W(\lambda), M(\lambda))$ and $\mathbf{S}' = \lambda \mathbb{T}(W_1, M_1) + (1-\lambda)\mathbb{T}(W_2, M_2)$, both have the same W, but $\mathbf{S}$ has a larger M. Because both $\mathbf{S}$ and $\mathbf{S}'$ are in the region where $\widetilde{J}^i_M \geq 0$, we have $\widetilde{J}^i(\mathbf{S}) \geq \widetilde{J}^i\left(\mathbf{S}'\right)$. Therefore,

$$\begin{aligned} J^i(W(\lambda), M(\lambda)) &= \widetilde{J}^i(\mathbb{T}(W(\lambda), M(\lambda))) \geq \widetilde{J}^i(\lambda\mathbb{T}(W_1, M_1) + (1-\lambda)\mathbb{T}(W_2, M_2)) \\ &\geq \lambda \widetilde{J}^i(\mathbb{T}(W_1, M_1)) + (1-\lambda)\widetilde{J}^i(\mathbb{T}(W_2, M_2)) \\ &= \lambda J^i(W_1, M_1) + (1-\lambda) J^i(W_2, M_2). \end{aligned}$$

It is easy to check that the resulting $J^i(W, M)$ ($j^i(w, m)$) satisfies all properties stated in Proposition 4 (Lemma 4). For completeness, we provide several properties of j^i on the domain above the curve $M^*(W)$. Notice that

$$j^i(w, m) = J^i(W, M) = \widetilde{J}^i\left(W, M^*(W)\right) = \widetilde{j}^i\left(W - \frac{U\left(m^*\right)}{r}, m^*\right),$$

where $m^* = M^* > M$. By construction, $J^i_M(W, M) = \frac{1}{\gamma r} j^i_w(w, m) + j^i_m(w, m) = 0$. Then, utilizing the fact that $\widetilde{J}^i_M\left(W, M^*(W)\right) = 0$ (therefore the indirect impact on m^* (or M^*) is zero), one can easily verify that

$$j^i_w(w, m) = \widetilde{j}^i_w\left(W - \frac{U\left(m^*\right)}{r}, m^*\right)$$

$$j^i_m(w, m) = -\frac{1}{\gamma r} j^i_w(w, m) = \widetilde{j}^i_m\left(W - \frac{U\left(m^*\right)}{r}, m^*\right) \geq \frac{1}{\gamma r m^*} \in \left(\frac{1}{\gamma r m}, \frac{1}{\gamma r \gamma_L}\right]$$

$$\frac{1}{\gamma r} j^i_{ww} + j^i_{wm} = J^i_{WM}(W, M) = 0, \text{ and } j^i_{mm} = \frac{1}{\gamma^2 r^2} j^i_{ww}.$$

A.6.3 Convergence and the upper-first-best states Let $\mathcal{C}(X)$ as the set of continuous, bounded, and concave functions on the convex compact set

$$X = \left\{(W, M) : M \in [\gamma_L, \gamma], W \in \left[\frac{U(M)}{r}, \frac{U(M)+b}{r}\right]\right\} \subset \mathbb{R}^2.$$

We have defined an operator $\mathbb{O} : \mathcal{C}(X) \to \mathcal{C}(X)$ to construct $J^i = \mathbb{O}\left(J^{i-1}\right)$ successively. Specifically, for $J^{i-1} \in \mathcal{C}(X)$, define J^i in two steps. First,

$$\widetilde{J}^i(W, M) = [b - rW + U(M)]^{1+\frac{p}{r}} \left[\int_{\frac{U(M)}{r}}^{W} \frac{pY - c(M) + pJ^{i-1}\left(x + \frac{b}{p}, M\right)}{[b - rx + U(M)]^{-2-\frac{p}{r}}}\right.$$

$$\left.\times dx + J^L\left(\frac{U(M)}{r}\right) b^{-1-\frac{p}{r}}\right] = \frac{rJ^L\left(\frac{U(m)}{r}\right)}{r+p} + \frac{p[b - rW + U(M)]^{1+\frac{p}{r}}}{r+p}$$

$$\times \left[\int_{\frac{U(M)}{r}}^{W} \frac{(r+p)\left(Y + J^{i-1}\left(x + \frac{b}{p}, M\right)\right)}{[b - rx + U(M)]^{-2-\frac{p}{r}}} dx + \frac{J^{i-1}\left(\frac{U(M)}{r}, M\right)}{b^{1+\frac{p}{r}}}\right],$$

since $J^{i-1}\left(\frac{U(M)}{r}, M\right) = J^L\left(\frac{U(M)}{r}\right)$. Second, the transformation $\mathbb{T}(W, M)$ defined in Equation (A10) gives $J^i(W, M) = \widetilde{J}^i(\mathbb{T}(W, M))$. Now we show that mapping $\mathbb{O}$ satisfies Blackwell's sufficient conditions for a contraction mapping (Stokey and Lucas 1989), which implies that there exists a unique J such that J^i converges to $J \in \mathcal{C}(X)$ uniformly.

We need to verify the monotonicity condition,

$$\mathbb{O}(f) \le \mathbb{O}(g) \text{ if } f \le g,\ f, g \in \mathcal{C}(X),$$

and the discounting condition,

$$\mathbb{O}(f + x) \le \mathbb{O}f + \frac{p}{r+p}x \text{ where } f \in \mathcal{C}(X), x \in \mathbb{R}.$$

To see the monotonicity condition, decompose $\mathbb{O}$ into $\mathbb{O}_1$ (from J^{i-1} to $\widetilde{J}^i$) and $\mathbb{O}_2$ (from $\widetilde{J}^i$ to J^i). If $f \le g$, then $\mathbb{O}_1 f \le \mathbb{O}_1 g$. Fix W, and let M_f^* and M_g^* be the corresponding compensation-setting curves. Clearly, if $M < \min\left(M_f^*, M_g^*\right)$, then $\mathbb{O}_2 f \le \mathbb{O}_2 g$ holds. If $M > \max\left(M_f^*, M_g^*\right)$,

$$\mathbb{O}_2(f)(W, M) = \mathbb{O}_1(f)\left(W, M_f^*\right) \le \mathbb{O}_1(g)\left(W, M_f^*\right) \le \mathbb{O}_2(g)\left(W, M_g^*\right) = \mathbb{O}_2(g)(W, M).$$

Finally, consider that M sits between M_f^* and M_g^*. Without loss of generality, consider $M_f^* < M_g^*$. Then,

$$\mathbb{O}_2(f)(W, M) = \mathbb{O}_1(f)\left(W, M_f^*\right) \le \mathbb{O}_1(g)\left(W, M_f^*\right) \le \mathbb{O}_1(g)(W, M) = \mathbb{O}_2(g)(W, M),$$

where the third inequality uses the fact that $\mathbb{O}_1(g)$ is concave, $M_f^* < M < M_g^*$, and M_g^* attains the maximum. The second discounting condition is straightforward.

Note that we have focused on the case $M > \gamma_L$; however, the previous construction also applies to the line with $M = \gamma_L$ and $W < \frac{U(\gamma_L)+b}{r}$. To complete the construction of J, we derive the value function for the upper-first-best states where $M = \gamma_L$ and $W \geq \frac{U(\gamma_L)+b}{r}$. Since the agent is risk-neutral, one particular solution has the agent consume $\frac{1}{\gamma_L}\left(W - \frac{U(\gamma_L)+b}{r}\right)$ whenever $W \geq \frac{U(\gamma_L)+b}{r}$; afterward, the state-pair stays at $\left(\frac{U(\gamma_L)+b}{r}, \gamma_L\right)$ without jumps, and the agent obtains $\frac{b}{p\gamma_L} < Y$ whenever a cash flow occurs. Based on Equation (23), it is easy to show that in this region

$$J(W, \gamma_L) = J\left(\frac{U(\gamma_L)+b}{r}, \gamma_L\right) - \frac{1}{\gamma_L}\left(W - \frac{U(\gamma_L)+b}{r}\right) = \frac{pY}{r} - \frac{u^{-1}(rW)}{r},$$

which is the first-best result when K, the maximum number of cash flows generated by the agent, is ∞. When K is finite, we can just replace $\frac{pY}{r}$ with $\frac{pY}{r}\left[1 - \left(\frac{p}{r+p}\right)^K\right]$ in the above equation.

A.7 Proof of Proposition 5

For any contract Π that satisfies the necessary conditions stated in Proposition 2, we introduce the investors' auxiliary gain process $G_t(\Pi)$ as

$$G_t(\Pi) = -\int_0^t e^{-rs} c_{s-} ds + \int_0^t e^{-rs} Y dN_s + e^{-rt} J(W_t, M_t). \tag{A11}$$

Recall the dynamics of two state variables in Equations (13), (14), (15), and (16):

$$dW_t = rW_{t-}dt - u(c_{t-})dt + \frac{b}{p}(dN_t - pdt),$$
$$dM_t = dM_t^D + \beta_t^M dN_t, \text{ where } dM_t^D \leq 0 \text{ and } dM_t^D \leq -\beta_t^M pdt,$$

where the relevant controls are dM_t^D and β_t^M.

For any incentive-compatible and no-savings contract Π, the investors' expected instantaneous gain $e^{rt}dG_t$ is

$$\mathbb{E}_{t-}\left[e^{rt}dG_t\right] = \left[\begin{array}{c} -rJ(W,M) - c(M) + p\left(Y + \left[J\left(W + \frac{b}{p}, M + \beta_t^M\right) - J(W,M)\right]\right) \\ +J_W \cdot (rW - U(M) - b) \end{array}\right] dt$$
$$+ \left[J\left(W, M + dM_t^D\right) - J(W,M)\right].$$

Note that $W = W_{t-}$ and $M = M_{t-}$. In the proof of Proposition 5, we show that the optimal policy to maximize $\mathbb{E}_{t-}\left[e^{rt}dG_t\right]$ is setting $dM_t^D = 0$ as in Equation (21), and $\beta_t^M = \min\left(M^*\left(W_{t-} + \frac{b}{p}\right) - M_{t-}, 0\right)$ as in Equation (27). Due to the construction in Section 4, we have $\mathbb{E}_{t-}\left[e^{rt}dG_t\right] = 0$ under the optimal policy, and $\mathbb{E}_{t-}\left[e^{rt}dG_t\right] \leq 0$ for other incentive-compatible and no-savings contracts. Then the standard verification argument leads to the following proposition. Finally, since J is concave, randomization cannot improve the investors' value.

The existence of $J(W, M)$ is established in Section 8. To maximize $\mathbb{E}_{t-}\left[e^{rt}dG_t\right]$, we need to maximize $pJ\left(W + \frac{b}{p}, M + \beta_t^M\right)dt + J\left(W, M + dM_t^D\right)$ (note that $W = W_{t-}$ and $M = M_{t-}$). Since $J_M(W, M) \geq 0$, it is without loss of generality to consider two cases: (i) $\beta_t^M \leq 0$ and $dM_t^D = 0$, and (ii) $\beta_t^M > 0$ and $dM_t^D = -\beta_t^M pdt$. We want to rule out the second case. If it is true, then $J\left(W, M + dM_t^D\right) = -J_M(W, M)\beta_t^M pdt$, and we are maximizing a function $B\left(\beta_t^M\right)$ such that

$$B\left(\beta_t^M\right) \equiv J\left(W + \frac{b}{p}, M + \beta_t^M\right) - J_M(W, M)\beta_t^M pdt.$$

It is easy to show that $B''\left(\beta_t^M\right) \leq 0$; then, since

$$B'\left(\beta_t^M\right)\Big|_{\beta_t^M=0} = J_W\left(W+\frac{b}{p}, M\right) - J_M(W,M) < 0,$$

$B\left(\beta_t^M\right)$ is maximized at $\beta_t^M = 0$, in contradiction to the second case. Therefore, we have shown that the first case holds; that is, $dM_t^D = 0$ and $\beta_t^M \leq 0$. Because $J_M(W,M) = 0$ for $M > M^*(W)$, the optimal $\beta_t^M = \min\left(M^*\left(W_{t-}+\frac{b}{p}\right) - M_{t-}, 0\right)$.

Now we show that the optimal policy solves the relaxed problem. Our road map is to show that $\mathbb{E}[G_\tau(\Pi)]$, which is the investors' value from any contract Π, has an upper bound $G_0 = J(W_0, M_0)$, that is, $\mathbb{E}[G_\tau(\Pi)] \leq G_0$; however, under the optimal contract Π^* with policy $dM_t^D = 0$ and $\beta_t^M = \min\left(M^*\left(W_{t-}+\frac{b}{p}\right) - M_{t-}, 0\right)$, $\mathbb{E}\left[G_\tau\left(\Pi^*\right)\right] = G_0$.

Given any contract Π that satisfies the necessary conditions to be incentive-compatible and no-savings, we can write the increment of gain process as

$$dG_t(\Pi) = \mu_G(t)\,dt + e^{-rt}\left[J\left(W+\frac{b}{p}, M+\beta_t^M\right) - J(W,M)\right](dN_t - p\,dt),$$

where one can easily check that due to construction, $\mu_{G(\Pi^*)}(t) = 0$ under the optimal policy, and $\mu_{G(\Pi)}(t) \leq 0$ for other contracts that satisfy the necessary conditions. And, clearly, $J\left(W+\frac{b}{p}, M+\beta_t^M\right) - J(W,M)$ is bounded (note that β_t^M is bounded as M is bounded; even in the first-best region where W_t might be unbounded, J is linear in W so $J\left(W+\frac{b}{p}, M+\beta_t^M\right) - J(W,M)$ is bounded); therefore,

$$\left\{\int_0^t e^{-rs}\left[J\left(W_{s-}+\frac{b}{p}, M_{s-}+\beta_s^M\right) - J(W_{s-}, M_{s-})\right](dN_s - p\,ds)\right\}$$

forms a well-defined martingale for $0 \leq t < \infty$. Because at the termination $J(W_\tau, M_\tau) = -F_\tau$, $\mathbb{E}[G_\tau(\Pi)]$ is the investors' payoff. Therefore, for any t,

$$\begin{aligned}\mathbb{E}[G_\tau(\Pi)] &= \mathbb{E}\left[G_{t\wedge\tau}\left(\widetilde{\Pi}\right) + \mathbf{1}_{t\leq\tau}\left[\int_t^\tau e^{-rs}(Y\,dN_s - c_s)\,ds - e^{-r\tau}F_\tau\right]\right] \\ &\leq G_0 + e^{-rt}\mathbb{E}\left[\int_t^\infty e^{-r(s-t)} Y\,dN_s\right]. \end{aligned} \tag{A12}$$

where $\mathbb{E}\left[\int_t^\infty e^{-r(s-t)} Y\,dN_s\right]$ represents the present value of firm's total cash flow (without early termination), which is bounded. Therefore, when $t \to \infty$, $\mathbb{E}[G_\tau(\Pi)] \leq G_0$ for any contract, while under the optimal contract with policy $dM_t^D = 0$ and $\beta_t^M = \min\left(M^*\left(W_{t-}+\frac{b}{p}\right) - M_{t-}, 0\right)$, $\mu_{G(\Pi^*)}(t) = 0$ implies that the inequality in (A12) holds in equality, and therefore $\mathbb{E}[G_\tau(\Pi)] = G_0$. This proves our claim.

A.7.1 Proof of Theorem 5 The proof is essentially the combination of Proposition 5 and the argument right before Theorem 5. The first-best result directly follows from Section 4.5. We have the following lemma to show formally that under the downward-rigid wage contract the agent is optimal to work and consume the wage.

Lemma 6. Suppose that the agent has a hypothetical saving of $S_0 \geq 0$. Then the agent's optimal value when facing the downward-rigid wage contract with a state-variable pair (W, M) is

$$V(M, W, S) = W - \Phi(M, 0) + \Phi(M, S), \tag{A13}$$

where

$$\Phi(M,S) = \frac{1-e^{-\gamma(c(M)+rS)}}{r} = \frac{1}{r} - \frac{M}{\gamma r}e^{-\gamma rS}.$$

Proof. Simple algebra yields $V_M(M,W,S) = \frac{1}{\gamma r}\left(1-e^{-\gamma rS}\right) \geq 0$, and $V_S(M,W,S) = Me^{-\gamma rS}$. Introduce the agent's auxiliary gain process as

$$G_t^A = \int_0^t e^{-rs}\left(U\left(\widehat{M}_s\right) + b\left(1-\frac{\widehat{a}_s}{p}\right)\right)ds + e^{-rt}V\left(M_{t-}, W_{t-}, S_t\right),$$

where the evolutions of state variables are

$$dS = rSdt + c(M)\,dt - c\left(\widehat{M}\right)dt,$$
$$dM = \beta_t^M dN_t,$$
$$dW = (rW - U(M) - b)\,dt + \frac{b}{p}dN_t(\widehat{a}),$$

and we use the actual marginal utility $\widehat{M}_s$ as the agent's control variable. It does not make a difference by using S_t or S_{t-} because S has continuous paths. Then,

$$\mathbb{E}_{t-}\left[e^{rt}dG_t^A\right] = U\left(\widehat{M}\right)dt + b\left(1-\frac{\widehat{a}}{p}\right)dt - rVdt + dW + V_M(M,W,S)\,dM + V_S(M,W,S)\,dS.$$

It is easy to see that $\widehat{a} = 0$ maximizes $\mathbb{E}_t\left[e^{rt}dG_t^A\right]$ (and strictly so when $\beta_t^M < 0$ and $S > 0$; when $S = 0$, $\widehat{a} = p$ is also optimal—this is the optimal policy along the equilibrium path). Then we have

$$\frac{\mathbb{E}_{t-}\left[e^{rt}dG_t^A\right]}{dt} \leq U\left(\widehat{M}\right) + rW - U(M) - rV + Me^{-\gamma rS}\left(rS + c(M) - c\left(\widehat{M}\right)\right).$$

The FOC of $\widehat{M}$ (recall the definition of $U(\cdot)$ in Equation (17) and $c(\cdot)$ in Equation (18)) yields that (we can also easily check that FOC is sufficient because RHS is concave in $\widehat{M}$)

$$\widehat{M} = Me^{-\gamma rS}.$$

Plugging in, we have $\mathbb{E}_{t-}\left[e^{rt}dG_t^A\right]/dt \leq 0$. Because the inequality could hold in equality when the optimal policy is used, a standard verification argument similar to the proof of Proposition 5 shows our claim. ■

Given this lemma, the agent's value without saving $S_0 = 0$ is just W, which is achieved by working and not saving. This proves our claim.

A.8 Appendix for Section 4

When a certain action a_t is implemented at time t, the evolution of W follows:

$$dW_t = rW_{t-}dt - U\left(M_{t-}\right)dt + \frac{b}{p}(a_t - p)\,dt + \beta_t^W\left(dN_t(a_t) - a_t dt\right),$$

where $\beta_t^W \leq (\geq)\frac{b}{p}$ if $a_t = 0(\overline{p})$ (the proof will be similar to that of Proposition 1; see also Sannikov 2008).

A.8.1 Suboptimality of shirking Suppose that at time t shirking is implemented; then, we must have

$$dW_t = rW_{t-}dt - U\left(M_{t-}\right)dt + \frac{b}{p}(0-p)\,dt + \beta_t^W dN_t\,(a=0),$$

where $\beta_t^W \le \frac{b}{p}$. Because there is no success when shirking is implemented, $dN_t\,(a_t = 0) = 0$. Moreover, to prevent the agent from saving, we must have $dM_t = dM_t^D \le 0$. Because $J_M \ge 0$, it is optimal to set $dM_t^D = 0$, and

$$e^{rt}dG_t \le \left[-rJ - c(M) + J_W \cdot (rW - U(M) - b)\right]dt.$$

To ensure that $e^{rt}dG_t \le 0$, we use the construction of $\widetilde{J}$ in ODE (23). Since the same ODE holds for J as $J(W,M) = \widetilde{J}\left(W, \min(M^*(W), M)\right)$, we have

$$rJ = pY - c(M) + p\left[J\left(W + \frac{b}{p}, M\right) - J\right] + J_W\,(rW - U(M) - b).$$

Rearranging terms, we need the following condition to ensure that $e^{rt}dG_t \le 0$:

$$Y + J\left(W + \frac{b}{p}, M\right) - J(W,M) \ge 0 \text{ for all } (W,M).$$

Because J is concave in W, $J(W,M) - J\left(W + \frac{b}{p}, M\right) \ge -J_W(W,M)\,\frac{b}{p}$. Since property 1 in Proposition 4 implies that $\frac{1}{\gamma_L} \ge -J_W$, we have the sufficient condition $Y \ge \frac{b}{p\gamma_L}$. This condition is ensured by the parameter restriction in condition (24). In addition, the condition is also necessary to rule out shirking, because it is the standard condition for the suboptimality of shirking when the agent becomes risk-neutral in the upper-first-best states (where J is linear in W). Intuitively, for working to be optimal, the expected cash flow pY should be greater than the upper bound of the agent's equivalent "monetary" effort cost, which is b/γ_L when the agent becomes sufficiently wealthy.

A.8.2 Suboptimality of myopic actions When myopic action is implemented, there is a noncontractible loss Δ due to the myopic action. On the benefit side, the myopic action boosts the cash-flow intensity to $\overline{p}$. We envision that the gain $\epsilon \equiv \overline{p} - p$ is small. Are there any other gains by implementing the myopic action in this model?

The answer is yes. In Remark 4, we note that the binding incentive-compatibility constraint $\beta_t^W = \frac{b}{p}$ plays a key role in invoking the joint-deviation argument in Section 3.2. Now, when $\beta_t^W > \frac{b}{p}$, the agent's incentive-compatibility constraint is slack, and condition (11) no longer holds. In other words, under a high-powered incentive scheme, the optimal contract punishes shirking severely and therefore deters the agent's joint deviation of "shirking and saving." As a result, cutting the agent's compensation after his failure—which is potentially a value-improving policy because $J_M > 0$—becomes possible.

In this case, because the unidimensional variable M is no longer sufficient to capture the agent's private-saving incentives, it is difficult to pinpoint the exact contractual gain of adjusting M upward following failures. Fortunately, we can use the necessary (local) no-savings condition under the effort choice $a = \overline{p}$ to bound this benefit. We can write the evolution of M as

$$\begin{aligned} dM_t &= dM_t^D + \beta_t^M dN_t\,(a_t = \overline{p}) \\ &= dM_t^D - \beta_t^M \overline{p}dt + \beta_t^M\,(dN_t\,(a_t = \overline{p}) - \overline{p}dt). \end{aligned}$$

Then, the no-savings condition under $a_t = \overline{p}$ requires that $dM_t^D \leq -\beta_t^M \overline{p}dt$, as dM_t has a nonpositive drift (supermartingale). Because $M_{t-} + \beta_t^M \geq \gamma_L$ must hold, we have a more explicit bound on the increment dM_t^D:

$$dM_t^D \leq \left(M_{t-} - \gamma_L\right) \overline{p}dt. \tag{A14}$$

This bound will be useful in showing our result.

If we implement the myopic action $a = \overline{p}$ at t, then for some $\beta_t^W \geq \frac{b}{p}$ the evolution of W_t is

$$dW_t = rW_{t-}dt - U\left(M_{t-}\right)dt + \frac{b}{p}\left(\overline{p} - p\right)dt + \beta_t^W\left(dN_t\left(\overline{p}\right) - \overline{p}dt\right).$$

We need to show that the auxiliary gain process G in Equation (A11) has a negative drift once $\overline{p}$ is implemented. Recall that by implementing the myopic effort, investors suffer a noncontractible loss Δ. Therefore, we have (recall that $\overline{p} = p + \epsilon$)

$$\mathbb{E}_{t-}\left[e^{rt}dG_t\right] = \begin{bmatrix} -rJ - c(M) + \overline{p}\left(Y + J\left(W + \beta_t^W, M + \beta_t^M\right) - J(W, M)\right) \\ +J_W\left(rW - U(M) + \frac{b\epsilon}{p} - \beta_t^W \overline{p}\right) \end{bmatrix}$$
$$\times\ dt + J\left(W, M + dM_t^D\right) - J(W, M) - \Delta dt.$$

We want to give an upper-bound estimate for $\mathbb{E}_{t-}\left[e^{rt}dG_t\right]$ given the condition $dM_t^D \leq -\beta_t^M \overline{p}dt$ and $\beta_t^W \geq \frac{b}{p}$.

Similar to the first paragraph in the proof of Proposition 5 in Appendix 8, we can show that setting $dM_t^D = -\beta_t^M \overline{p}dt$ and choosing the lowest (most negative) β_t^M maximizes $\mathbb{E}_{t-}\left[e^{rt}dG_t\right]$. Because of Equation (A14), the lowest possible β_t^M is $\gamma_L - M_t \leq 0$. Therefore, we have

$$\mathbb{E}_{t-}\left[e^{rt}dG_t\right] \leq \begin{bmatrix} -rJ - c(M) + \overline{p}\left(Y + \left[J\left(W + \beta_t^W, \gamma_L\right) - J(W, M)\right]\right) \\ +J_W\left(rW - U(M) + \frac{b\epsilon}{p} - \beta_t^W \overline{p}\right) \end{bmatrix}$$
$$\times dt + J_M(W, M)\,\overline{p}\,(M - \gamma_L)\,dt - \Delta dt$$
$$\leq \begin{bmatrix} -rJ - c(M) + \overline{p}\left(Y + \left[J\left(W + \beta_t^W, M\right) - J(W, M)\right]\right) \\ +J_W\left(rW_t - U(M) + \frac{b\epsilon}{p} - \beta_t^W \overline{p}\right) \end{bmatrix}$$
$$\times dt + J_M(W, M)\,\overline{p}\,(M - \gamma_L)\,dt - \Delta dt,$$

where the second inequality is due to $J_M \geq 0$. Now, the only choice variable is β_t^W; because J is concave,

$$\max_{\beta_t^W \geq \frac{b}{p}} \overline{p}J\left(W + \beta_t^W, M\right) - J_W \beta_t^W \overline{p}$$

yields a solution of $\beta_t^W = \frac{b}{p}$. Therefore,

$$\mathbb{E}_{t-}\left[e^{rt}dG_t\right] \leq \begin{bmatrix} -rJ - c(M) + \overline{p}\left(Y + \left[J\left(W + \frac{b}{p}, M\right) - J\right]\right) \\ +J_W\left(rW - U(M) - b\right) \end{bmatrix}$$
$$\times\ dt + J_M(W, M)\,\overline{p}\,(M - \gamma_L)\,dt - \Delta dt$$
$$= \epsilon\left[Y + \left[J\left(W + \frac{b}{p}, M\right) - J\right] + J_M(W, M)\left(M_t - \gamma_L\right)\right]$$
$$\times dt\ + J_M(W, M)\,p\,(M - \gamma_L)\,dt - \Delta dt.$$

We take ϵ to be arbitrarily small. Because $J_{WM} < 0$, when M is fixed, J_M attains the maximum when $W = \frac{U(M)}{r}$. Therefore, a sufficient condition, which can be verified easily ex post, is

$$\Delta > \max_{M \in [\gamma_L, \gamma]} p(M - \gamma_L) J_M \left(\frac{U(M)}{r}, M \right). \tag{A15}$$

Because the actual gain (subject to additional constraints regarding the agent's other deviating strategies) must be smaller, we provide a sufficient condition for the suboptimality of implementing the myopic action.

A.8.3 Verifying the optimality of working Combining the above results, we have the following proposition.

Proposition 6. Under conditions (24) and (A15), it is always optimal to implement working.
Proof. Take the auxiliary gain process G_t as defined in Equation (A11). In Section A.8.1 and Section A.8.2, we have shown that whenever actions other than working are implemented,

$$dG_t = \mu_G(t)\, dt + e^{-rt} \left[J\left(W + \beta_t^W, M + \beta_t^M \right) - J(W, M) \right] (dN_t(a_t) - a_t dt),$$

where $\mu_G(t) \leq 0$. We require β_t^W to be bounded in any feasible contract; because β_t^M has to be bounded since M is bounded, then $J\left(W + \beta_t^W, M + \beta_t^M\right) - J(W, M)$ is bounded (even in the first-best region where W might be unbounded—see the argument in the proof of Proposition 5 in Appendix 8). Consequently,

$$\left\{ \int_0^t e^{-rs} \left[J\left(W_{s-} + \frac{b}{p}, M_{s-} + \beta_s^M \right) - J(W_{s-}, M_{s-}) \right] (dN_s - p ds) \right\}$$

forms a well-defined martingale for $0 \leq t < \infty$. We then can invoke the same argument as in the proof of Proposition 5 to show that the contract given in Theorem 5 (which implements always working) is optimal among all contracts that may implement other actions. ■

A.9 Appendix for Section 6.1

Following Sannikov (2008), we denote the investors' concave value function as $f(W)$, and continuation payoff W follows

$$dW = \left(rW_{t-} - u\left(c_{t-}^* \right) \right) dt + \frac{b}{p} (dN_t - pdt),$$

where c^* solves the investors' HJB equation

$$rf(W) = \max_{c \geq 0} \left\{ pY - c + p\left[f\left(W + \frac{b}{p} \right) - f(W) \right] + f'(W)[rW - u(c) - b] \right\}. \tag{A16}$$

Clearly, due to the risk-neutrality for a sufficiently high consumption level, similar to the previous discussion there is an absorbing first-best state for $W \geq \frac{U(\gamma_L)+b}{r}$, and $f'(W) = \frac{1}{\gamma_L}$. Note that in Sannikov (2008) the upper-absorbing state corresponds to the case where the wealth effect becomes extreme, and the firm is terminated. The difference is purely due to different utility specifications.

In the lower region where $f'(W) > -\frac{1}{\underline{\gamma}}$, it is easy to show that the optimal wage policy, as a function of W, is

$$c^* = \begin{cases} \frac{1}{\gamma} \ln\left(\frac{-1}{f'(W)\gamma}\right) & \text{when } f'(W) < -\frac{1}{\gamma} \\ 0 & \text{otherwise} \end{cases}.$$

This policy can be understood as follows. In Equation (A16), paying one more dollar of wage has a unit marginal cost, and on the benefit side, it reduces the agent's continuation payoff by $u'(c)$, so the marginal benefit is $-f'(W)\,u'(c)$. The above policy equates the marginal cost with the marginal benefit whenever possible. As f is concave, c^* will bind at zero for low W's, which reflects the fact that when the inefficient termination (once $W = 0$) is close, the marginal benefit of reducing continuation payoff $-f'(W)\,u'(c)$ either is small, or even becomes negative.

A.10 Appendix for Section 7.1

We again construct $J^{RP}(W, M)$ recursively. The following lemma lists the properties of $j^{RP,i-1}(W, M)$. In property 4, $\underline{w}^{i-1}(m)$ is the renegotiation curve discussed in the main text, and $W^{RP,i-1,*}(m)$ is the compensation-setting curve similar to the definition in Equation (25).

Lemma 7. For the compensation-setting stage value function $j^{RP,i-1}(w, m)$, we have the following properties:

1. $-\frac{1}{\gamma_L} \le j_w^{RP,i-1} \le 0$, $j_m^{RP,i-1} > \frac{1}{\gamma r m}$, and $0 \le \frac{1}{\gamma r} j_w^{RP,i-1} + j_m^{RP,i-1} \le \frac{1}{\gamma r m}$.
2. $j_{ww}^{RP,i-1} < 0$, $j_{mm}^{RP,i-1} < 0$, $j_{wm}^{RP,i-1} > 0$, and $j_{ww}^{RP,i-1} j_{mm}^{RP,i-1} - \left(j_{wm}^{RP,i-1}\right)^2 \ge 0$. Therefore, $j^{RP,i-1}(w, m)$ is concave.
3. $\frac{1}{\gamma r} j_{ww}^{RP,i-1} + j_{wm}^{RP,i-1} \le 0$, $\frac{1}{\gamma r} j_w^{RP,i-1}\left(\frac{b}{r}, m\right) + j_m^{RP,i-1}\left(\frac{b}{r}, m\right) \ge 0$, and $\frac{1}{\gamma r} j_w^{RP,i-1}\left(\frac{b}{r}, m\right) + j_m^{RP,i-1}\left(\frac{b}{r}, m\right) = 0$.
4. $\underline{w}^{i-1}(m) < W^{RP,i-1,*}(m) - \frac{U(m)}{r} \le \frac{b}{r}$, $\underline{w}^{i-1\prime}(m) \ge 0$.

Consider the production stage in the i^{th} subperiod. There exists a curve $\underline{w}^i(m)$ such that $\tilde{j}^{RP,i}$ takes the value $J^L\left(\frac{U(m)}{r}\right)$, and $\tilde{j}_w^{RP,i} = 0$ on this curve. Similar to Equation (A4), one can check that

$$\tilde{j}_w^{RP,i}(w, m) = p\,[b - rw]^{\frac{p}{r}} \left\{ \int_{\underline{w}^i(m)}^{w} j_w^{RP,i-1}\,[b - rx]^{-1-\frac{p}{r}}\,dx + \frac{\widehat{Y} + j^{RP,i-1}\left(\underline{w}^i(m) + \frac{b}{p}, m\right) - J^L\left(\frac{U(m)}{r}\right)}{\left(b - r\underline{w}^i(m)\right)^{\frac{p}{r}}} \right\},$$

where $\widehat{Y} \equiv \frac{pY - rL}{p}$. Due to renegotiation-proof, at $\underline{w}^i(m)$, $\tilde{j}_w^{RP,i} = p\left[Y + j^{RP,i-1}\left(\underline{w}^i(m) + \frac{b}{p}, m\right) - J^L\left(\frac{U(m)}{r}\right)\right] = 0$. Therefore, we define

$$\underline{w}^i(m) = \inf\left\{0 \le x \le \frac{b}{r} : \left[\widehat{Y} + j^{RP,i-1}\left(x + \frac{b}{p}, m\right) - J^L\left(\frac{U(m)}{r}\right)\right] = 0\right\}. \quad \text{(A17)}$$

We assume that $\underline{w}^i(m)$ defined in Equation (A17) satisfies $\underline{w}^i(m) < \frac{b}{r}$. Because j is decreasing in x, this condition holds when L is relatively large so that $\widehat{Y} = \frac{pY - rL}{p}$ is relatively small. Under this condition, we can show that $\underline{w}^i(m) < W^{RP,i-1,*}(m) - \frac{U(m)}{r}$. For instance, when $M = \gamma_L$, for $\underline{W}\left(\underline{\gamma}\right)$ to take a value below the compensation-setting point $W^*(\gamma_L) = \frac{U(\gamma_L)+b}{r}$, we require that the investors' value at termination is greater than their value at the upper-first-best boundary point; that is, $J^L\left(\frac{U(\gamma_L)}{r}\right) > J\left(\frac{U(\gamma_L)+b}{r}, \gamma_L\right) \Leftrightarrow rL - pY > -\frac{b}{\gamma_L} \Leftrightarrow \widehat{Y} < \frac{b}{p\gamma_L}$. We have the following lemma for $\widetilde{j}^{RP,i}$.

Lemma 8. For the production-stage value function $\widetilde{j}^{RP,i}(w, m)$, we have

1. $\widetilde{j}_w^{RP,i} < 0$, $\widetilde{j}_m^{RP,i} > \frac{1}{\gamma rm}$, and $\frac{1}{\gamma r}\widetilde{j}_w^{RP,i} + \widetilde{j}_m^{RP,i} \leq \frac{1}{\gamma rm}$.
2. $\widetilde{j}_{ww}^{RP,i} < 0$, $\widetilde{j}_{mm}^{RP,i} < 0$, $\widetilde{j}_{wm}^{RP,i} > 0$, and $\widetilde{j}_{ww}^{RP,i}\widetilde{j}_{mm}^{RP,i} - \left(\widetilde{j}_{wm}^{RP,i}\right)^2 > 0$. Therefore, $\widetilde{j}^{RP,i}(w, m)$ is concave.
3. $\frac{1}{\gamma r}\widetilde{j}_{ww}^{RP,i} + \widetilde{j}_{wm}^{RP,i} \leq 0$, $\frac{1}{\gamma r}\widetilde{j}_w^{RP,i}\left(\frac{b}{r}, m\right) + \widetilde{j}_m^{RP,i}\left(\frac{b}{r}, m\right) < 0$.
4. $\underline{w}^{i\prime}(m) \geq 0$.

For detailed proofs, see He (2008). When L is small (for instance, $L = 0$), $\underline{w}(m)$ and $W^{RP,*}(m) - \frac{U(m)}{r}$ both bind at $\frac{b}{r}$. At this point, without success the agent stays at that point, and after a jump the agent is promoted to another point with a lower m (higher wages). Because the termination is extremely inefficient ($pY > \frac{b}{\gamma_L}$, so keeping the project alive is always better off), termination will be off-equilibrium.

References

Allen, F. 1987. Repeated Principal-agent Relationships with Lending and Borrowing. *Economic Letters* 17:27–31.

Bebchuk, L., and J. Fried. 2004. *Pay Without Performance: The Unfulfilled Promise of Executive Compensation.* Cambridge, MA: Harvard University Press.

Berkovitch, E., R. Israel, and Y. Spiegel. 2000. Managerial Compensation and Capital Structure. *Journal of Economics & Management Strategy* 9:549–84.

Bettis, J., J. Bizjak, J. Coles, and S. Kalpathy. 2010. Stock and Option Grants with Performance-based Vesting Provisions. *Review of Financial Studies* 23:3849–88.

Biais, B., T. Mariotti, G. Plantin, and J. Rochet. 2007. Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications. *Review of Economic Studies* 74:345–90.

Bisin, A., and A. Rampini. 2006. Markets as Beneficial Constraints on the Government. *Journal of Public Economics* 90:601–29.

Bizer, D., and P. DeMarzo. 1999. Optimal Incentive Contracts When Agents Can Save, Borrow, and Default. *Journal of Financial Intermediation* 8:241–49.

Chen, C., J. Liang, and S. Lin. 2006. Unexpected Earnings, Abnormal Accruals, and Changes in CEO Bonuses. *International Journal of Accounting Studies* (special issue):25–50.

Cole, H., and N. Kocherlakota. 2001. Efficient Allocations with Hidden Income and Hidden Storage. *Review of Economic Studies* 68:523–42.

Cremers, M., and D. Palias. 2011. Tenure and CEO Pay. Working Paper, Yale University.

DeMarzo, P., and M. Fishman. 2007. Optimal Long-term Financial Contracting. *Review of Financial Studies* 20:2079–2128.

DeMarzo, P., M. Fishman, Z. He, and N. Wang. Forthcoming. Dynamic Agency and q Theory of Investment. *Journal of Finance*.

DeMarzo, P., and Y. Sannikov. 2006. Optimal Security Design and Dynamic Capital Structure in a Continuous-time Agency Model. *Journal of Finance* 61:2681–2724.

Doepke, M., and R. Townsend. 2006. Dynamic Mechanism Design with Hidden Income and Hidden Auctions. *Journal of Economic Theory* 126:235–85.

Dynan, K., W. Edelberg, and M. Palumbo. 2009. The Effects of Population Aging on the Relationship Among Aggregate Consumption, Saving, and Income. *American Economic Review: Papers & Proceedings* 99:380–86.

Efendi, J., A. Srivastava, and E. Swanson. 2007. Why Do Corporate Managers Misstate Financial Statements? The Role of In-the-money Options and Other Incentives. *Journal of Financial Economics* 85:667–708.

Eisfeldt, A., and A. Rampini. 2008. Managerial Incentives, Capital Reallocation, and the Business Cycle. *Journal of Financial Economics* 87:177–99.

Fernandes, A., and C. Phelan. 2000. A Recursive Formulation for Repeated Agency with History Dependence. *Journal of Economic Theory* 91:223–47.

Fudenberg, D., B. Holmstrom, and P. Milgrom. 1990. Short-term Contracts and Long-term Agency Relationships. *Journal of Economic Theory* 51:1–31.

Gompers, P., J. Ishii, and A. Metrick. 2003. Corporate Governance and Stock Prices. *Quarterly Journal of Economics* 118:107–55.

Harris, M., and B. Holmstrom. 1982. A Theory of Wage Dynamics. *Review of Economic Studies* 49:315–33.

Hart, O., and J. Moore. 1998. Default and Renegotiation: A Dynamic Model of Debt. *Quarterly Journal of Economics* 113:1–41.

Hart, O., and J. Tirole. 1988. Contract Renegotiation and Coasian Dynamics. *Review of Economic Studies* 55:509–40.

He, Z. 2008. A Continuous-time Contracting Problem with Private Savings. Ph.D. dissertation, Northwestern University.

He, Z. 2009. Optimal Executive Compensation When Firm Size Follows Geometric Brownian Motion. *Review of Financial Studies* 22:859–92.

———. 2011. A Model of Dynamic Compensation and Capital Structure. *Journal of Financial Economics* 100:351–66.

Holmstrom, B., and P. Milgrom. 1987. Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica* 55:303–28.

———. 1991. Multi-task Principal-agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization* 7:24–52.

Hopenhayn, H., and J. Nicolini. 1997. Optimal Unemployment Insurance. *Journal of Political Economy* 105:412–38.

Jensen, M. 2005. Agency Costs of Overvalued Equity. *Financial Management* 34:5–19.

Jenter, D., and K. Lewellen. 2010. Performance-induced CEO Turnover. Working Paper, Stanford University.

Kaplan, S., and B. Minton. 2008. How Has CEO Turnover Changed? Working Paper, University of Chicago.

Karatzas, I., and S. Shreve. 1988. *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag.

Kocherlakota, N. 2004. Figuring Out the Impact of Hidden Savings on Optimal Unemployment Insurance. *Review of Economic Dynamics* 7:541–54.

Larkin, I. 2006. The Cost of High-powered Incentives: Salesperson Gaming in Enterprise Software. Working Paper, Harvard Business School.

Levitt, S., and S. Dubner. 2005. *Freakeconomics: A Rogue Economist Explores the Hidden Side of Everything*. New York: Harper/Collins.

Maskin, E., and J. Tirole. 1999. Unforseen Contingencies and Incomplete Contracts. *Review of Economic Studies* 66:83–114.

Mitchell, M., and Y. Zhang. 2007. Unemployment Issuance and Hidden Savings. Working Paper, University of Iowa.

Murphy, K. 1999. *Executive Compensation*. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Ch. 38, pp. 2485–2563. Amsterdam: Elsevier/North Holland.

Phelan, C., and R. Townsend. 1991. Computing Multi-period, Information-constrained Optima. *Review of Economic Studies* 58:853–81.

Piskorski, T., and A. Tchistyi. 2010. Optimal Mortgage Design. *Review of Financial Studies* 23:3098–3140.

Protter, P. 1990. *Stochastic Integration and Differential Equations*. New York: Springer-Verlag.

Rogerson, W. 1985. Repeated Moral Hazard. *Econometrica* 53:69–76.

Rubinstein, M. 1995. On the Accounting Valuation of Employee Stock Options. *Journal of Derivatives* 3:8–24.

Sannikov, Y. 2008. A Continuous-time Version of the Principal-agent Problem. *Review of Economic Studies* 75:957–84.

Spear, S., and S. Srivastava. 1987. On Repeated Moral Hazard with Discounting. *Review of Economic Studies* 54:599–607.

Spear, S., and C. Wang. 2005. When to Fire a CEO: Optimal Termination in Dynamic Contracts. *Journal of Economic Theory* 120:239–56.

Stein, J. 1989. Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior. *Quarterly Journal of Economics* 14:655–69.

Stokey, N., and R. Lucas. 1989. *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press.

Werning, I. 2001. Repeated Moral-hazard with Unmonitored Wealth: A Recursive First-order Approach. Working Paper, Massachusetts Institute of Technology.

Williams, N. 2006. On Dynamic Principal-agent Problems in Continuous Time. Working Paper, University of Wisconsin–Madison.

Yermack, D. 2006. Golden Handshakes: Separation Pay for Retired and Dismissed CEOs. *Journal of Accounting and Economics* 41:237–56.