

# Only time will tell: A theory of deferred compensation and its regulation\*

Florian Hoffmann<sup>†</sup>      Roman Inderst<sup>‡</sup>      Marcus Opp<sup>§</sup>

December 2016

## Abstract

We characterize optimal contracts in settings where the principal observes informative signals over time about the agent's one-time action. If both are risk-neutral contract relevant features of any signal process can be represented by a deterministic informativeness process that is increasing over time. The duration of pay trades off the gain in informativeness with the costs resulting from the agent's liquidity needs. The duration is shorter if the agent's outside option is higher, but may be non-monotonic in the implemented effort level. We evaluate effects of regulatory proposals that mandate the deferral of bonus payments and use of clawback clauses.

*Keywords:* Compensation design, duration of pay, short-termism, moral hazard, persistence, principal-agent models, informativeness principle, financial regulation.

*JEL Classification:* D86 (Economics of Contract: Theory), G28 (Government Policy and Regulation), G21 (Banks, Depository Institutions, Mortgages).

---

\*The paper subsumes insights from an earlier working paper titled “Regulating deferred incentive pay.” We thank Willie Fuchs, Nicolae Gârleanu, Ben Hermalin, John Kuong, Robert Marquez, John Morgan, Christine Parlour, Steve Tadelis, Jean Tirole, and Jeffrey Zwiebel for valuable insights. In addition, we thank seminar participants at Berkeley (Micro Theory), MIT (Organizational Econ), Harvard (Organizational Econ), Harvard Business School, University of Bonn (Micro Theory), University of Mannheim (Applied Micro), University of British Columbia, University of Calgary, INSEAD, EFA Oslo, CEPR Gerzensee. Hoffmann and Inderst gratefully acknowledge financial support from the ERC (Advanced Grant “Regulating Retail Finance”).

<sup>†</sup>University of Bonn. E-mail: fhoffmann@uni-bonn.de.

<sup>‡</sup>Johann Wolfgang Goethe University Frankfurt. E-mail: nderst@finance.uni-frankfurt.de.

<sup>§</sup>Harvard University (on leave from UC Berkeley). E-mail: marcus.opp@fas.harvard.edu.

# 1 Introduction

In many real-life principal-agent relationships, actions by agents have long-lasting – *not immediately observable* – effects on outcomes. For example, within the financial sector, investments by private equity fund managers only produce verifiable returns to investors upon an exit, a credit rating issued by a credit rating agency (or a loan officer’s loan decision) can be evaluated more accurately over the lifetime of a loan, and, a bank’s risk management is only stress-tested in times of crisis. Outside the financial sector, innovation activities by researchers, be it in academia or in industry, typically produce signals such as patents or citations only with considerable delay. Similarly, the quality of a CEO’s strategic decisions may not be assessed until well into the future. The list is certainly not exclusive and, yet, it suggests that the delay of observability is an important, if not defining feature of many moral hazard environments.

The goal of our paper is to address a basic research question. How does one optimally structure the *intertemporal* provision of incentives in “*only-time-will-tell*” information environments when the agent’s liquidity needs make it costly to defer compensation? The contribution of our paper vis-à-vis the existing literature is to allow for abstract, general information systems, and, yet, obtain a tractable characterization of optimal contracts that can be readily applied to many economic settings.

In a first key step, we show that, for any signal process, the information content relevant for the optimal timing of pay is fully summarized by the maximal likelihood ratio across signal histories, which is an increasing function of time. It measures by how much the principal is able to reduce agency rents by deferring longer. The optimal duration of pay trades off this benefit of deferral with the costs resulting from the agent’s liquidity needs. We find that the duration of pay is decreasing in the agent’s outside option. Surprisingly, the duration may be non-monotonic in the induced action. The latter comparative statics depend on the signal process, in particular on whether the principal “learns” faster under a high or low action.

While our focus is abstract and theoretical in nature, understanding the basic forces behind the duration of pay is of applied interest. Researchers (Bebchuk and Fried (2010)) and regulators (e.g., Carney (2014)) alike have argued that compensation packages for CEOs are too “*short-termist*.” This diagnosis has motivated regulators across the world to intervene in the structure of compensation packages and impose minimum deferral requirements and clawback clauses in the financial sector.<sup>1</sup> For example, in the United

---

<sup>1</sup>In the EU, a new directive adopted in 2010 includes strict rules for bank executives’ bonuses. Directive 2010/76/EU, amending the Capital Requirements Directives, which took effect in January 2011. It has already been fully implemented in a number of countries, including France, Germany, and

Kingdom, the regulator imposes a minimum deferral period of 3 years for bonuses to executives in the financial sector. Our framework lends itself to analyze the effect of such policies on the action that the principal, the board of a company, induces in equilibrium. We show that “moderate” deferral requirements nudge the principal to induce more “long-term” actions. Deferral regulation effectively works as a (Pigouvian) tax on “short-term” actions: Actions that the principal would optimally implement with short-term payout dates become more costly. However, we also show that sufficiently long minimum deferral periods unambiguously backfire.

We derive these results in a parsimonious modeling framework in order to focus on the optimal intertemporal provision of incentives in general “*only-time-will-tell*” information environments. That is, we analyze a setting with a one-time action (such as Holmstrom (1979) or Harris and Raviv (1979)) with bilateral risk-neutrality and limited liability on the side of the agent (cf. Innes (1990) and Kim (1997)). The agent chooses an unobservable, continuous action that affects the distribution of a process of contractible signals, such as output realizations, defaults, annual performance reviews, etc. These signals may arrive continuously or at discrete points in time. A compensation contract stipulates (bonus) payments to the agent, conditioning on all information available at a particular date (the history of signals), and must satisfy both the agent’s incentive compatibility and participation constraint. Finally, as in DeMarzo and Duffie (1999), we model the agent’s liquidity needs via relative impatience.

Following the structure of Grossman and Hart (1983), our paper first characterizes compensation contracts that minimize compensation costs to implement a given action. The key simplification of this problem results from constructing a quantitative measure of the gain in informativeness over time, formally capturing “*only time will tell.*” The intuitive construction draws on well-known insights from static models (Innes (1990)) and adapts them to our dynamic information environment. We initially show that optimal contracts typically fall into the class of *maximal-incentives* contracts. If a maximal-incentives contract stipulates a bonus at some date  $t$ , then this bonus is only paid for an outcome (here, a history of realized signals up to date  $t$ ) that maximizes the likelihood ratio across all possible date- $t$  histories. Without a risk-sharing motive, it is optimal to punish the agent for all other histories, i.e., pay out zero due to limited liability of the agent. Then, by tracing out this maximal date- $t$  likelihood ratio over time, we obtain an increasing function of time, the informativeness function. This single-valued function fully encodes all features of the signal process relevant for the duration of pay.

The duration of pay, the weighted average payout time, is then determined by a simple

---

the UK and has led to mandatory deferral of bonuses for several years.

cost-benefit trade-off. The benefit of deferral is captured by the informativeness function. The cost of deferral results from the deadweight costs due to relative impatience. The optimal intertemporal provision of incentives then critically depends on the size of the agent's outside option. When the agent's outside option is low, the duration of pay reflects a rent-extraction motive. More informative signals allow the principal to reduce the agency rent, and the principal will optimally defer as long as the growth rate of informativeness exceeds the growth rate of impatience costs, as given by the difference in discount rates. In turn, when the agent's outside option is sufficiently high, the size of the compensation package is fixed, so the principal's rent extraction motive is absent. Then, optimal contracts simply minimize weighted average impatience costs subject to ensuring incentive compatibility. We show that this results in at max two payout dates regardless of the underlying signal process.

This general characterization allows us to analyze the comparative statics of contract duration and, hence, the degree of *short-termism* in optimal contracts. The duration is shorter if the agent's outside option is higher, but the effect of the induced effort level on contract duration is ambiguous. The intuition is that contracts provide incentives via two substitutable channels, the size of the agent's compensation package and contract duration. When the (binding) outside option exogenously increases the size of the compensation package, the principal optimally responds by reducing the contract duration. In contrast, while higher effort requires more incentives, it is generally unclear whether the principal provides these additional incentives optimally via higher pay, a higher duration, or both channels. Intuitively, the optimal adjustment along all margins depends on whether informativeness grows faster for high or low effort, which depends on the underlying signal process.

The just characterized optimal contract solves the first problem of Grossman and Hart (1983) and determines the minimum wage cost to implement a given action, or simply, the wage cost function. Then, the induced equilibrium action solves the second problem of Grossman and Hart (1983) by trading off the benefits of an action, the present value of gross profits accruing to the principal, with the associated wage cost.

In the context of the financial sector, many reasons have been put forward for why this privately chosen action is not socially optimal, either because the benefit of an action to the bank's board, the principal, does not account for an externality (cf. Allen and Gale (2000)) or because it reflects a corporate governance problem (cf. Kuhn and Zwiebel (2009)). Hence, there is scope for regulatory intervention. We are, thus, interested in how the induced action is affected by recent regulatory interventions in the financial sector, in particular, minimum deferral periods. That is, our goal is a comparative statics analysis

of the action choice in a general agency model, rather than the broader goal of optimal regulation. An advantage of our narrower objective is that it allows us to remain agnostic about the source of the externality that motivates the regulator to intervene.

How does deferral regulation operate? Minimum deferral regulation *only* affects the (compensation) cost side, but does not affect the benefit side of the principal’s problem.<sup>2</sup> If regulation constrains the principal in his preferred timing choice for a given action, it forces him to adjust the compensation contract along other dimensions, such as the size or contingency of pay, to be able to implement a given action. By optimality of the original compensation package, these adjustments must strictly increase the level of compensation cost, akin to a tax. Deferral thus “works” if it makes the action that society prefers over the laissez-faire outcome differentially less costly, and sufficiently so to induce an action change; i.e., if it taxes the bad, “short-term action” differentially more than the good, “long-term” action.

It is easiest to illustrate the “tax” intuition with a stylized example. Consider a discrete action set  $\mathcal{A} = \{a_0, a_1, a_2\}$  and the case where the agent has no outside option. Here,  $a_0$  refers to the agent’s zero-cost action which the principal can induce at zero compensation costs. Let  $a_1$  refer to the action that the principal incentivizes in the absence of regulation. Finally,  $a_2$  corresponds to the action preferred by society with  $a_2 \succ a_1 \succ a_0$ . To capture the regulator’s diagnosis that observed compensation contracts are “too short-termist,” we posit that the pay duration corresponding to the respective actions satisfies  $T(a_0) < T(a_1) < T(a_2)$  under the optimal contract without regulation. Now consider the effect of imposing a minimum deferral period of  $T(a_2)$ . By construction, this minimum deferral period constrains the optimal implementation for actions  $a_0$  and  $a_1$ , but does not affect the cost of inducing action  $a_2$ . Thus, the good news is that under deferral regulation, the differential wage cost of implementing  $a_2$  relative to  $a_1$  is *reduced*. If this pushes the differential cost below the (unaffected) differential benefit, then regulation “works!” However, the bad news is that deferral simultaneously *increases* the differential wage cost of  $a_1$  relative to  $a_0$ . Implementing action  $a_1$  requires a rent which can no longer be paid in the most cost-efficient way since regulation binds,  $T(a_1) < T(a_2)$ ; costs go up. In contrast, while regulation also binds for action  $a_0$ , the costs of implementing action  $a_0$  are unaffected as  $a_0$  requires no incentive pay: A forced deferral of zero pay is not costly.<sup>3</sup> The general insight is that the tax induced by deferral is not only lower (*ceteris paribus*) for long-term actions — where regulation does not bind —

---

<sup>2</sup>This contrasts, for instance, with higher capital requirements that do not affect the wage cost function, but force shareholders to internalize more of the downside risk.

<sup>3</sup>The argument extends to the case where the (marginal) cost of action  $a_0$  is sufficiently small.

but also for actions that require little or no incentive pay to begin with, as impatience costs associated with deferral are increasing in the size of pay. Importantly, this insight is robust regardless of whether the action is interpreted as effort or risk-taking or a combination thereof. In our stylized example, the tax burden is, thus, non-monotonic in the regulator’s action ranking, i.e., zero for  $a_0$ , positive for  $a_1$ , and zero for  $a_2$ . This non-monotonicity of the tax burden, induced by the interaction with the size of pay, is responsible for why deferral regulation can only have limited effects and, eventually, backfires. If action  $a_2$  is sufficiently close to action  $a_1$ , in terms of the corresponding *net* profits in the absence of regulation, then deferral regulation has the desired effect of *nudging* the principal to induce the long-term action  $a_2$  by sufficiently raising the cost of action  $a_1$ . If  $a_2$  is sufficiently unattractive to the principal to begin with, deferral risks backfiring as action  $a_1$  unambiguously becomes more costly relative to  $a_0$ . The latter case always applies as the minimum deferral period becomes sufficiently large.

On an abstract level, the limited effectiveness of deferral regulation is due to the fact that it only targets a symptom, the “short-termist” compensation package, rather than the cause of shareholders’ preference for the “short-term” action, such as bailout guarantees or corporate governance problems. By intervening in the timing of pay the regulator still leaves the principal room to adjust other relevant margins of the compensation contract according to his (biased) preferences. If the regulator wants to achieve more far-reaching effects purely by intervening in compensation design, it would need to eliminate the possibility of “contracting around regulation” by dictating all terms of the compensation contract. In this vein, we show that regulatory malus requirements, i.e., clawbacks of promised, but yet unpaid bonuses from an escrow account, are one step in this direction as they further restrict the contracting space. Of course, such proposals require a high degree of sophistication by regulators, a concern already voiced in the literature on Pigouvian taxation (see e.g., Baumol (1972)).

**Literature.** The premise of our paper is that the timing of pay determines the information about the agent’s hidden action that the principal can use for incentive compensation. This relates our analysis to the broader literature on comparing information systems in agency problems, which derives sufficient conditions for information to have value for the principal. In particular, implementation costs are shown to be lower for an information system that is “more informative” (Holmstrom (1979)), Blackwell sufficient (Gjesdal (1982) and Grossman and Hart (1983)) or has a likelihood ratio distribution that is dominated in the sense of second-order stochastic dominance (Kim (1995)). Time generates a family of information systems via the arrival of additional signals, such that

a principal must do (weakly) better when having access to an information system generated by a later date. In a risk-neutral setting like ours, the measurement of “how much better” is exactly determined by the increase in the *maximal* likelihood-ratio.<sup>4</sup> The key difference of our paper relative to this classical strand of the literature is that having access to a better information system generates (endogenous) costs due to the agent’s liquidity needs. The optimal timing of pay resolves the resulting trade-off, thereby determining the optimal information system in equilibrium (see Theorem 1). Viewed from this angle, minimum deferral regulation can then be interpreted as an intervention that “forces” the principal to obtain more information, on which he can then condition incentive pay. However, since the associated costs are not exogenously fixed, but interact with the agency problem, this regulation may have non-trivial effects on the equilibrium action choice.

More concretely, our paper belongs to a small, but growing literature that analyzes moral hazard setups where actions have persistent effects. Most closely related is Hopenhayn and Jarque (2010) who consider the case of a risk-averse agent. Due to the agent’s desire to smooth consumption across states (and time),<sup>5</sup> risk-aversion generates different trade-offs in compensation design, such that the entire likelihood ratio distribution matters to capture the benefit of deferral in their setup. This makes it difficult to sharply characterize the optimal duration of pay even for the special case of i.i.d. signals with binary outcomes. Our model with an impatient but risk neutral agent allows for a complete characterization of the optimal timing of pay for general (discrete as well as continuous) signal processes. We thus nest the important finance application of moral hazard by a securitizer of defaultable assets, as studied in Hartman-Glaser, Piskorski, and Tchisty (2012) and Malamud, Rui, and Whinston (2013).

In our setup, the only role of the timing of pay is to improve the information system available to the principal. As is well known, in repeated-action settings the timing of pay may play an important role even when actions are immediately and perfectly observed (cf. Ray (2002)): In this literature, backloading of rewards to the agent has the benefit that it incentivizes both current as well as future actions.<sup>6</sup> Work by Jarque (2010), Sannikov (2014), or Zhu (2016) combines the effects of repeated actions and persistence. The additional complexity, however, requires special assumptions on the signal process.

---

<sup>4</sup> Cf. also Chaigneau, Edmans, and Gottlieb (2016) for an option pricing approach to quantifying the cost savings associated with “more precise” information in optimal contracting.

<sup>5</sup> With Epstein-Zin preferences, it would be possible to separate the elasticity of intertemporal substitution and risk-aversion.

<sup>6</sup> Opp and Zhu (2015) analyze a repeated-action setting, where these incentive benefits derived from backloading have to be traded off against the agent’s liquidity needs.

Instead, our setup tries to isolate one effect, the idea that information gets better over time, and studies it in (full) generality.

**Organization of paper.** Our paper is organized as follows. In Section 2 we present the model setup and characterize optimal contracts to implement a given action. In Section 3, we characterize the equilibrium action choice and how it is affected by deferral regulation. Section 4 concludes. Proofs not contained in the main text are relegated to Appendix A. Appendix B contains supplementary material.

## 2 Optimal compensation design

### 2.1 Setup

We consider a principal-agent problem with bilateral risk-neutrality and limited liability of the agent in which the principal observes informative signals about the agent’s action over time. In this section, we will focus on optimal compensation design, i.e., characterize cost-minimizing contracts to implement a given action (the first problem in Grossman and Hart (1983)). We discuss the principal’s preferences over actions and the resulting equilibrium action choice, the second problem in Grossman and Hart (1983), in Section 3.

Time is continuous  $t \in [0, \bar{T}]$ . At time 0, an agent  $A$  with outside option  $v$  takes an unobservable action  $a \in \mathcal{A} = [0, \bar{a}]$  at cost  $c(a)$ . The cost function  $c(a)$  satisfies the usual conditions, i.e., it is strictly increasing and strictly convex with  $c(0) = c'(0) = 0$  as well as  $c'(\bar{a}) = \infty$ . The one-time action  $a$  affects the distribution of a stochastic process of verifiable signals  $X_t$  that may arrive continuously or at discrete points in time.<sup>7</sup> These abstract signals may correspond to output realizations, annual performance reviews by the principal, or, more generally, any (multidimensional) combination of informative signals. The respective date- $t$  history of realized signals  $h^t = \{x_j\}_{0 \leq j \leq t} \in H^t$  then constitutes the relevant performance measure as it captures the entire information about the agent’s action available to the principal at time  $t$ . Thus, it is useful to take a reduced form approach and define the respective probability measures directly over histories. For each given  $(a, t)$ , let  $\mu_t(\cdot|a)$  denote the parameterized probability measure on the set of histories  $H^t$ . Then, for any subset  $\tilde{H}^t \subset H^t$ ,  $\mu_t(\tilde{H}^t|\cdot)$  maps  $\mathcal{A}$  into  $\mathbb{R}$  such that  $\mu_t(H^t|a) = 1$ . We impose the standard Cramér-Rao regularity conditions used in statistical inference (cf., e.g., Casella and Berger (2002)): First, for each  $t$ , the set  $H^t$  is independent of the agent’s action choice, i.e., no shifting support. Second, the log

---

<sup>7</sup> Formally, the index set of the stochastic process  $X_t$  can be any subset of  $[0, \bar{T}]$ .

likelihood ratio (score),  $\frac{d \log L_t}{da}$ , exists and is bounded for any  $t$  and  $h^t$ . Here, by convention of the statistics literature the likelihood function,  $L_t(a|h^t)$ , refers to the density if history  $h^t$  has zero mass, while  $L_t(a|h^t) = \mu_t(\{h^t\}|a)$  otherwise.

The following two information environments are stylized examples of this framework and will be used throughout the text to illustrate our notation and general findings.

**Example 1** *Discrete information arrival: At each  $t \in \{1, 2, \dots, \bar{T}\}$  there is an i.i.d. binary signal  $x_t \in \{s, f\}$  where the probability of success “s” is given by  $a \in [0, 1]$ .*

**Example 2** *Continuous information arrival:  $X_t$  is a (stopped) counting process where  $x_t = 1$  indicates that failure has occurred before time  $t$  ( $x_t = 0$  otherwise). The action  $a$  affects the survival function  $S(t|a)$  such that the failure rate  $\lambda(t|a) = -\frac{d \log S(t|a)}{dt}$  is strictly decreasing in  $a$ ,  $\forall t \in (0, \bar{T})$ ,  $a$ .<sup>8</sup>*

Thus, in Example 1 the probability measure associated with a history of two successes satisfies  $\mu_2(\{(s, s)\}|a) = L_2(a|(s, s)) = a^2$ . In Example 2, the date- $t$  history without previous failure is summarized by the last signal,  $x_t = 0$ , and has positive probability mass, i.e.,  $\mu_t(\{x_t = 0\}|a) = L_t(a|x_t = 0) = S(t|a)$ . In contrast, a history  $\tilde{h}_t$  with a failure at any particular date  $t$  has zero mass, so  $L_t(a|\tilde{h}_t) = S(t|a)\lambda(t|a)$ .

A compensation contract  $\mathcal{C}$  stipulates transfers from the principal to the agent as a function of the information available at the time of payout. We assume that the principal is able to commit to long-term contracts. Formally, such a contract is represented by the cumulative compensation process  $b_t$  (progressively measurable with respect to the filtration generated by  $X_t$ ). In particular,  $db_t$  denotes the instantaneous bonus paid out to the agent at time  $t$ . Limited liability of the agent implies that  $b_t$  must be non-decreasing. While deferring payments allows the principal to condition bonuses on more informative signals, it is costly to do so since the agent has liquidity needs. In particular, as is standard in dynamic principal-agent models, we make the assumption that the agent is relatively impatient.<sup>9</sup> The discount rates of the agent,  $r_A$ , and the principal,  $r_P$ , satisfy

$$\Delta_r := r_A - r_P > 0.$$

We now state the principal’s compensation design problem for a given action  $a \in (0, \bar{a})$ , i.e., we solve for the contract  $\mathcal{C}$  that implements  $a$  at the lowest present value of wage cost (discounted at the principal’s rate of time preference  $r_P$ ), denoted as  $W(a)$ .

<sup>8</sup>The assumption that the hazard rate is strictly decreasing in  $a$  plays a similar role as the monotone likelihood ratio property (MLRP) in static principal-agent models with immediately observable signals, see, e.g., Milgrom (1981). See Appendix B.3 for concrete parametric examples.

<sup>9</sup>See, e.g., DeMarzo and Duffie (1999), DeMarzo and Sannikov (2006), or Opp and Zhu (2015).

**Problem 1 (Compensation design)**

$$W(a) := \min_{b_t} \mathbb{E} \left[ \int_0^{\bar{T}} e^{-r_P t} db_t \middle| a \right] \quad s.t. \quad (1)$$

$$V_A := \mathbb{E} \left[ \int_0^{\bar{T}} e^{-r_A t} db_t \middle| a \right] - c(a) \geq v \quad (\text{PC})$$

$$a = \arg \max_{\tilde{a} \in \mathcal{A}} \mathbb{E} \left[ \int_0^{\bar{T}} e^{-r_A t} db_t \middle| \tilde{a} \right] - c(\tilde{a}) \quad (\text{IC})$$

$$db_t \geq 0 \quad \forall t \quad (\text{LL})$$

The first constraint is the agent's time-0 participation constraint (PC).<sup>10</sup> The agent's utility  $V_A$ , i.e., the present value of compensation discounted at the agent's rate net of the cost of the action, must at least match her outside option  $v$ . Second, incentive compatibility (IC) requires that it is optimal for the agent to choose action  $a$  given  $\mathcal{C}$ . Limited liability of the agent (LL) imposes a lower bound on the transfer to the agent, i.e.,  $db_t \geq 0$ . For ease of exposition, we initially neither impose an upper bound on transfers (e.g., limited liability of the principal) nor a monotonicity constraint on the principal's payoff function (cf. Innes (1990)).<sup>11</sup>

As is common in static moral hazard problems with continuous actions (see e.g., Holmstrom (1979) and Shavell (1979)) we assume that the first-order approach is valid. Hence, for each  $a$ , we replace (IC) by the following first-order condition

$$\frac{\partial}{\partial a} \mathbb{E} \left[ \int_0^{\bar{T}} e^{-r_A t} db_t \middle| a \right] = c'(a). \quad (2)$$

As is well known, this only requires that the necessary condition for agent optimality in (2) is also sufficient *at the optimal contract*. We thus provide sufficient conditions for the validity of the first-order approach after we present the optimal contract in Theorem 1 (cf., in static settings, Rogerson (1985) and Jewitt (1988)). Here, it is useful to point out that our characterization of the optimal contract readily extends to settings with a binary action set where no such additional technical conditions have to be imposed (see

<sup>10</sup>Since the agent in our model only chooses an action once at time 0 and is protected by limited liability, the participation constraint of the agent only needs to be satisfied at  $t = 0$ .

<sup>11</sup>The monotonicity constraint in Innes (1990) requires payments to the agent to be non-decreasing in the principal's profits, which gives rise to the optimality of debt contracts. By imposing such a constraint in our framework it would be possible to endogenize the optimal choice of debt maturity.

Appendix B.1).

## 2.2 Analysis

### 2.2.1 Maximal-incentives contracts

Our preliminary goal is to show that a contract solving Problem 1 typically falls in the class of “maximal-incentives” contracts. The construction of this contract class draws on well-known insights from static moral hazard models. In static principal-agent models with bilateral risk-neutrality and limited liability of the agent maximal-incentives contracts are optimal as long as there is a relevant incentive constraint:<sup>12</sup> Due to the absence of risk-sharing considerations, the agent is only rewarded for the outcome which is most indicative of the recommended action, i.e., the outcome with the highest likelihood ratio given this action (and obtains zero for all other outcomes due to limited liability).

To extend the definition of these maximal-incentives contracts to our dynamic setting, we first define the most informative outcome for a given  $t$ , i.e., the *most informative history* using date- $t$  information. As informativeness is measured in likelihood ratio units, let  $h_{MI}^t(a)$  denote the history that maximizes the log likelihood ratio over all possible date- $t$  histories  $H^t$ . Of course, as in a static setting, i.e., the special case of our model with  $\bar{T} = 0$ , we need to impose that such a maximizer exists.<sup>13</sup> Then, we have

$$h_{MI}^t(a) := \arg \max_{h^t \in H^t} \frac{d \log L_t(a|h^t)}{da}.^{14} \quad (3)$$

Concretely, in Example 1 the most informative histories are success in  $t = 1$ , i.e.,  $h_{MI}^1(a) = (s)$  and a sequence of successes by date 2, i.e.,  $h_{MI}^2(a) = (s, s)$ , etc. For our Example 2, the assumption on the hazard rate implies that the most informative history at each date  $t$ ,  $h_{MI}^t(a)$ , is survival up to  $t$ , which is summarized by  $x_t = 0$ . In general,  $h_{MI}^t(a)$  might well change with  $a$  and  $h_{MI}^{t+1}(a)$  need not be a continuation history of  $h_{MI}^t(a)$  (see Appendix B.2 for an illustration based on a non-i.i.d. version of Example 1 in the spirit of Manso (2011)). In any case, paying a date- $t$  reward following the realization of history  $h_{MI}^t(a)$  provides the strongest incentives per unit of expected pay among

---

<sup>12</sup>See e.g., Innes (1990) or, for a textbook treatment, Tirole (2005). If the incentive constraint is not relevant for compensation costs, the compensation design problem generically has a multiplicity of solutions. One possible class of contracts which allows to minimize compensation costs is the class of bonus contracts (see Kim (1997)).

<sup>13</sup>To see the technical issue, suppose that the time-0 signal  $x_0$  is drawn from a probability distribution without compact support, such as the normal distribution. Moreover, suppose that the likelihood ratio for a given action is increasing in the signal realization of  $x_0$  (such as under MLRP), then  $h_{MI}^0(a)$  does not even exist in this static environment.

<sup>14</sup>Uniqueness of this maximizer is not essential. It is without loss of generality to pick any maximizer.

all performance measures available at time  $t$ . Then, bilateral risk neutrality implies the optimality of maximal-incentives contracts:

**Lemma 1** *As long as the incentive problem is relevant for compensation costs, i.e., the shadow price on (IC) satisfies  $\kappa_{IC} > 0$ , the optimal contract never stipulates rewards for any history other than  $h_{MI}^t$  histories. That is, for all  $t$ ,  $db_t = 0$  whenever  $h^t \neq h_{MI}^t(a)$ .*

Going forward, we use the shorthand notation  $\mathcal{C}_{MI}$ -contracts for this class of contracts. To provide intuition for the proof, it is useful to introduce the concept of the *size of the compensation package*, which is defined as the agent's time-0 valuation of the compensation package (using her discount rate)

$$B := \mathbb{E} \left[ \int_0^{\bar{T}} e^{-r_A t} db_t \middle| a \right]. \quad (4)$$

Then, (PC) requires that  $B \geq v + c(a)$ . Now, the only reason for why the principal's wage costs in an optimal contract,  $W(a)$ , may exceed the minimum wage cost imposed by (PC),  $v + c(a)$ , is a relevant incentive constraint. In the interesting case when (IC) is relevant, the agent, thus, either obtains an agency rent,  $B > v + c(a)$ , and/or the principal defers some payouts beyond date 0 resulting in a wedge between  $W$  and  $B$  due to agent impatience. It is now easy to see that contracts outside the class of  $\mathcal{C}_{MI}$ -contracts are strictly suboptimal, i.e., yield strictly higher wage costs to the principal. By shifting rewards towards  $h_{MI}^t$  histories, the principal would either be able to reduce the size of the compensation package  $B$  or move payments to an earlier date (or both) while preserving incentive compatibility and satisfying (PC).

When the (IC) constraint is irrelevant for compensation costs, the principal can achieve the minimum wage cost imposed by (PC),  $W = B = v + c(a)$ , by making all payments at time 0 (cf. the proof of Lemma 1). Since it is thus only interesting to analyze optimal payout times when (IC) is relevant, we will initially focus on this case (leading to Theorem 1) before characterizing optimal contracts for the alternative, less interesting case in Lemma 3.

### 2.2.2 Optimal payout times

The principal's choice of payout times  $t$  can be interpreted as choosing the quality of the information system. This optimization over the timing dimension is absent in static settings. Specifically, in Example 1 we have to determine whether the principal should optimally stipulate rewards for dates  $0, 1, \dots, \bar{T}$  or any combination of these, while in Example 2 the choice is from  $t \in [0, \bar{T}]$ . Intuitively, optimal payout times are pinned down

by the trade-off between *impatience costs*, measured by  $e^{\Delta r t}$ , and gains in *informativeness*. The key simplification of our analysis results from the fact that  $\mathcal{C}_{MI}$ -contracts only stipulate rewards for maximally informative histories. Without a risk-sharing motive, it is possible to disregard informativeness of performance signals associated with any non- $h_{MI}^t$  history. Date- $t$  informativeness is then appropriately measured by the maximal likelihood ratio at date  $t$ .

**Proposition 1** *Maximal informativeness  $I(t|a)$  is an increasing function of time*

$$I(t|a) := \max_{h^t \in H^t} \frac{d \log L_t(a|h^t)}{da} = \frac{d \log L_t(a|h_{MI}^t)}{da}. \quad (5)$$

The proof shows that starting from any possible history it is always possible to find a continuation path along which the likelihood ratio (weakly) increases. Formally, this follows from the fact that the likelihood ratio (score),  $\frac{d \log L_t(a|h^t)}{da}$ , is a martingale. Intuitively, since the principal can observe the entire history of signals he could always choose to ignore additional signals if he wanted to do so. Thus,  $I(t|a)$  must be an increasing function of time. It is a deterministic function, as it already maximizes over all possible realizations of histories  $H^t$ .<sup>15</sup>  $I(t|a)$  is bounded by the Cramér-Rao regularity conditions.

To illustrate how the characteristics of the signal process  $X$  map into (maximal) informativeness  $I$ , it is useful to revisit our two examples. In the discrete Example 1, the informativeness function is a right-continuous step function with  $I(t|a) = \frac{t}{a}$  for  $t = 1, 2, \dots, \bar{T}$ . Intuitively, since a “success” outcome is more informative if it is less likely to occur,  $I(t|a)$  is decreasing in  $a$ . In the continuous Example 2, survival up to time  $t$  is the most informative history, so that  $I(t|a) = \frac{d \log S(t|a)}{da}$ . Since the probability mass of failure in each instant is zero, the absence of failure at time 0 is not informative, i.e.,  $S(0|a) = 1$  for all  $a$  and, hence,  $I(0|a) = 0$ . Moreover, informativeness grows at a faster rate, as measured by  $\partial I(t|a) / \partial t$  if the hazard rate is more sensitive to the action, i.e.,  $\frac{\partial I(t|a)}{\partial t} = -\frac{\partial \lambda(t|a)}{\partial a} > 0 \quad \forall t \in (0, \infty), a$ . To see the intuition for this, think of approximating the local incentive constraint with a binary action choice. If the hazard rate under the high and low action is identical at time  $t$ , the principal learns nothing from the absence (or occurrence) of failure. If instead, the hazard rate under the low action is much higher than under the high action, the principal learns “a lot.” In Appendix B.3, we illustrate the informativeness functions  $I$  for several commonly used survival distributions (see Appendix-Figure 6). In particular, for a Poisson process with arrival rate  $\lambda(t|a) = \frac{1}{a}$ ,

<sup>15</sup> It is possible to relate  $I(t|a)$  to the well known Fisher information function in statistics, which for a given  $t$ , is defined as the variance of the likelihood ratio (score) across histories. The relevant notion of informativeness in our agency setting is the maximum score across histories as defined in (5).

we obtain that the informativeness function,  $I(t|a) = \frac{t}{a^2}$ , is linear in  $t$ , as for any i.i.d. signal process (cf. also Example 1 above).<sup>16</sup>

To determine the optimal timing of  $\mathcal{C}_{MI}$ -contracts it is now convenient to define  $w_s$  as the (expected) fraction of the compensation package that the agent derives from stipulated payouts up to time  $s$ , i.e.,

$$w_s := \mathbb{E} \left[ \int_0^s e^{-r_A t} db_t \middle| a \right] / B, \quad (6)$$

so that  $w_{\bar{T}} = \int_0^{\bar{T}} dw_t = 1$ . Hence,  $\int_0^{\bar{T}} t dw_t$  measures the *duration of the compensation contract*.<sup>17</sup> Then, Problem 1 can be written as a deterministic problem by viewing  $B$  and  $w$  as control variables (rather than  $b$ ).<sup>18</sup>

**Problem 1\***

$$W(a) = \min_{B, w_t} B \int_0^{\bar{T}} e^{\Delta r t} dw_t$$

$$B \geq v + c(a) \quad (\text{PC})$$

$$B \int_0^{\bar{T}} I(t|a) dw_t = c'(a) \quad (\text{IC})$$

$$dw_t \geq 0 \quad \forall t \text{ and } w_{\bar{T}} = 1 \quad (\text{LL})$$

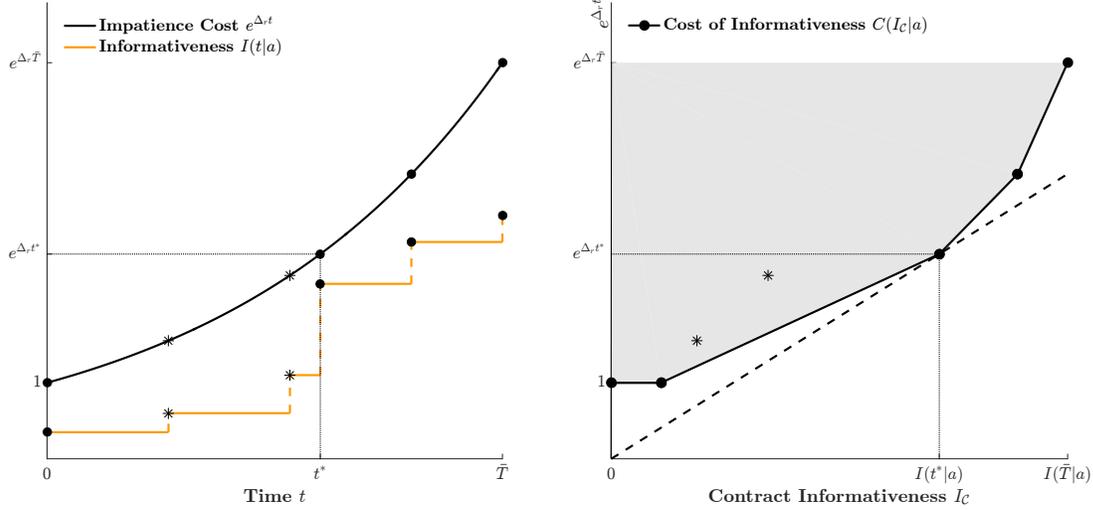
The transformed problem clearly shows that as long as two signal generating processes give rise to the same maximal informativeness,  $I(t|a)$ , the timing of pay in the respective optimal contracts will be identical (even if likelihood ratios differ along less informative histories). Thus, when (IC) is relevant, one may think of the function  $I(t|a)$  (rather than  $X_t$ ) as the primitive of the information environment that fully captures the formalization of “time will tell.” This benefit of deferral, in the sense of a more informative performance measure, however, comes at a cost due to the agent’s relative impatience.

More concretely, from Problem 1\*, the optimal timing of pay for  $\mathcal{C}_{MI}$ -contracts is determined by the trade-off between weighted average *impatience costs*,  $\int_0^{\bar{T}} e^{\Delta r t} dw_t$ , and the weighted average *informativeness*,  $\int_0^{\bar{T}} I(t|a) dw_t$ , or simply, *contract informativeness*

<sup>16</sup> As will become evident below, the discrete Example 1 with  $I(t|a) = \frac{t}{a}$  and the continuous Example 2 for the case of a Poisson process with  $I(t|a) = \frac{t}{a^2}$  will produce identical payout times for  $a = 1$  (abstracting from integer problems).

<sup>17</sup> We thus employ a duration measure analogous to the Macaulay duration which is standard in the fixed-income literature; the weights of each payout date are determined by the present value of the associated payment divided by the size of the compensation package.

<sup>18</sup> Given  $B$ ,  $w_t$  and  $h_{MI}^t(a)$ , one may readily solve for  $db_t$  from  $\mathbb{E}[db_t|a]e^{-r_A t} = Bdw_t$ . Of course, this requires that a solution to Problem 1 exist (see discussion in Section 2.2.4).



**Figure 1. Impatience costs versus informativeness:** The left panel plots both impatience costs,  $e^{\Delta r t}$ , and maximal informativeness,  $I(t|a)$ , as a function of time  $t$  for an example economy that features a significant increase in informativeness at date  $t^*$ . The right panel plots weighted average impatience costs against weighted average informativeness  $I_{\mathcal{C}}$  for all possible contracts. The solid circles in the right panel correspond to the boundary points that define the (minimum) cost of informativeness  $C(I_{\mathcal{C}}|a)$ .

$I_{\mathcal{C}}$ . To extend the notion of contract informativeness beyond the class of  $\mathcal{C}_{MI}$ -contracts, we define

$$I_{\mathcal{C}} := \frac{d \log B}{da},$$

which captures the contract incentives per unit of expected pay. Its units can formally be interpreted as a weighted average log likelihood ratio of all performance signals used in the compensation contract  $\mathcal{C}$ . Intuitively, if a contract only stipulates unconditional payments at every date, then  $I_{\mathcal{C}} = 0$ , since the compensation package does not vary with  $a$ . If the contract is a  $\mathcal{C}_{MI}$ -contract, then  $I_{\mathcal{C}} = \int_0^{\bar{T}} I(t|a) dw_t$ . The shaded region in the right panel of Figure 1 depicts *all feasible* combinations of average impatience costs and contract informativeness  $I_{\mathcal{C}}$  corresponding to the informativeness and impatience cost function plotted in the left panel. It may be constructed as the convex hull of  $\mathcal{C}_{MI}$ -contracts with single payout dates  $\{(I(t|a), e^{\Delta r t})\}_{t \in [0, \bar{T}]}$ , and, in addition, two uninformative contracts stipulating unconditional payments at date 0 and  $\bar{T}$  corresponding to points  $(0, 1)$  and  $(0, e^{\Delta r \bar{T}})$ . Of particular relevance for the subsequent analysis is the lower boundary  $C(I_{\mathcal{C}}|a)$  of the convex hull.

**Definition 1** *The cost of informativeness,  $C(I_{\mathcal{C}}|a)$ , is the minimal impatience cost for a given level of contract informativeness  $I_{\mathcal{C}}$ .*

For the example information environment in Figure 1, the cost of informativeness is depicted by the solid black line in the right panel. By construction,  $C$  is an increasing and convex function mapping  $I_{\mathcal{E}} \in [0, I(\bar{T}|a)]$  into  $[1, e^{\Delta_r \bar{T}}]$ .  $\mathcal{C}_{MI}$ -contracts play an essential role for this construction, since for all  $I_{\mathcal{E}} \geq I(0|a)$ , the lower boundary,  $C$ , can only be achieved by using  $\mathcal{C}_{MI}$ -contracts. Moreover, since it is always possible to use a  $\mathcal{C}_{MI}$ -contract that exclusively stipulates payments at date  $t$ ,  $C(I(t|a)|a)$  is bounded above by  $e^{\Delta_r t}$ , i.e.,  $C(I(t|a)|a) \leq e^{\Delta_r t} \quad \forall t$ .

For ease of exposition, our subsequent analysis of the timing of pay considers optimal payout dates for the case when (PC) is slack and when (PC) binds separately. Theorem 1 then synthesizes these results and provides conditions for when each case applies.

**PC slack.** First, consider the case when (PC) is slack. Then, the optimal timing reflects the principal's rent extraction motive: The principal can reduce the size of the agent's compensation package by deferring longer and hence using more informative performance signals. However, deferral does not imply a zero-sum transfer of surplus to the principal, but instead involves deadweight costs due to relative impatience. A (generically) unique payout time optimally resolves this trade-off

$$T_{RE}(a) = \arg \min_t \frac{e^{\Delta_r t}}{I(t|a)}, \quad (7)$$

which implies contract informativeness of  $I_{\mathcal{E}} = I(T^*(a)|a)$ . Graphically, it can be identified by the point on  $(I_{\mathcal{E}}, C(I_{\mathcal{E}}|a))$  that minimizes the slope of a ray through the origin.<sup>20</sup> From the right panel of Figure 1, we can immediately deduce that  $T_{RE}(a) = t^*$  in this example. Moreover, if  $I$  is differentiable in  $t$ , as is e.g., the case in Example 2, we can characterize  $T_{RE}(a)$  in terms of an intuitive first-order condition:

$$\left. \frac{d \log I(t|a)}{dt} \right|_{t=T_{RE}(a)} = \Delta_r. \quad (8)$$

That is, the principal defers until the (log) growth rate of informativeness,  $\frac{d \log I}{dt}$ , equals the (log) growth rate of impatience costs,  $\Delta_r$ .

**PC binds.** In contrast, when (PC) binds the size of the agent's compensation package is fixed at  $B = v + c(a)$ , so that the principal's rent extraction motive is absent. The

<sup>19</sup>In the knife-edge case of multiple global minimizers, one can use Pareto optimality as a criterion to select the earliest payout date. The agent strictly prefers the one with the earliest payout date since  $V_A = \frac{c'(a)}{I(T_{RE}|a)} - c(a)$ , while the principal is indifferent.

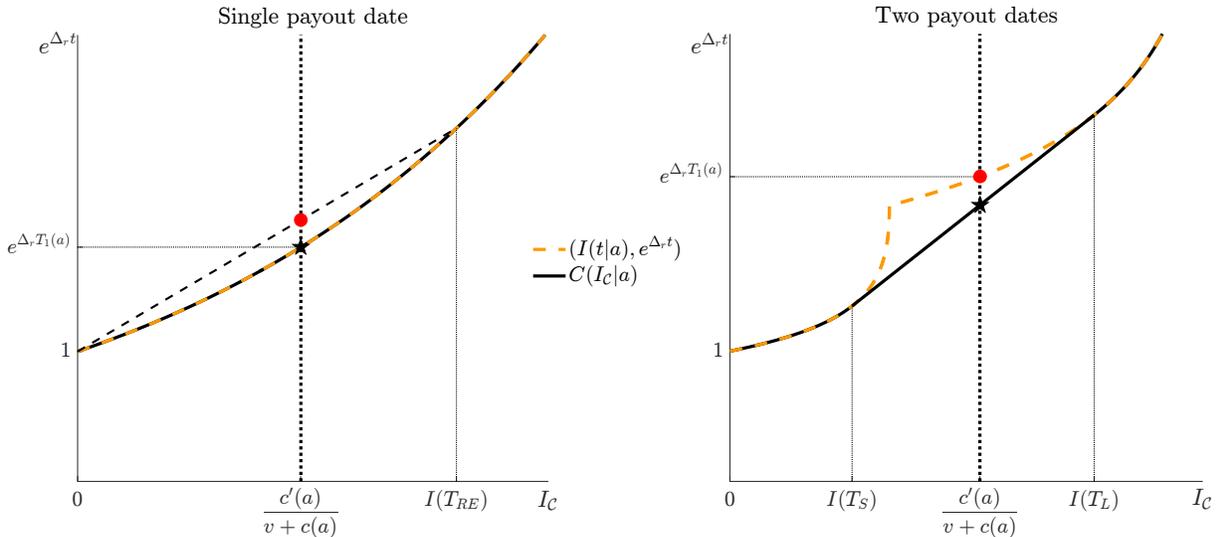
<sup>20</sup>This minimum slope can be interpreted in economic terms as the shadow price on (IC),  $\kappa_{IC}$ .

choice of payout times minimizes the contract's impatience costs,  $\int_0^{\bar{T}} e^{\Delta r t} dw_t$ , subject to ensuring incentive compatibility which requires from (IC) that contract informativeness satisfies  $I_{\mathcal{C}} = \int_0^{\bar{T}} I(t|a) dw_t = \frac{c'(a)}{v+c(a)}$ . Since minimal impatience costs  $C\left(\frac{c'(a)}{v+c(a)} \mid a\right)$  can be achieved by using a convex combination of at most two payout dates (see right panel of Figure 1), this puts a tight upper bound on the number of payout dates specified by optimal contracts with binding (PC).

**Lemma 2** *Timing of  $\mathcal{C}_{MI}$ -contracts with binding (PC)*

**Single-date:** *The optimal payout occurs at a single date if and only if there is a date  $T_1(a)$  that solves  $I(T_1|a) = \frac{c'(a)}{v+c(a)}$  and, given a solution, this payout date achieves minimal impatience costs, i.e.,  $e^{\Delta r T_1(a)} = C\left(\frac{c'(a)}{v+c(a)} \mid a\right)$ .*

**Two-dates:** *Otherwise, the contract requires a short-term payout date  $T_S$  and a long-term date  $T_L$ . Then,  $(I(T_S|a), e^{\Delta r T_S})$  and  $(I(T_L|a), e^{\Delta r T_L})$  are defined as the respective boundary points whose convex combination defines  $C$  for  $I(T_S|a) < I_{\mathcal{C}} < I(T_L|a)$ .<sup>21</sup>*



**Figure 2. PC binds one vs. two dates:** The left panel plots the case when  $C$  is a strictly convex function (so that a given level of informativeness is optimally achieved with one payout date). In the right panel, single payout dates are strictly suboptimal for any informativeness level satisfying  $I_{\mathcal{C}} \in (I(T_S|a), I(T_L|a))$ .

Lemma 2 implies that there are generically two payment dates for discrete information processes, as there is typically no single date  $T_1(a)$  that solves  $I(T_1|a) = \frac{c'(a)}{v+c(a)}$ . For

<sup>21</sup> In general, these two payout dates are unique (as in Figure 2). In some knife-edge cases, there may be multiple solutions for  $T_S$  and  $T_L$ . In such a knife-edge case, it is without loss of generality to pick any 2 of these payout dates.

illustrative purposes, it is thus instructive to explain the economics behind the choice of one versus two payment dates in an environment where  $I$  is continuous as is, e.g., the case in Example 2.<sup>22</sup> In the left panel of Figure 2,  $C$  is strictly convex so any convex combination of two payment dates is inefficient. One may have conjectured that the optimal contract with binding (PC) can be decomposed into the optimal contract under slack (PC), which pays out at date  $T_{RE}(a)$ , and an additional sufficiently high (unconditional) date-0 payment to satisfy (PC). This conjecture is typically wrong. The candidate contract (indicated by the red circle) turns out to produce strictly higher wage costs to the principal than the contract with a single payout date at  $T_1(a)$  (indicated by the black star).

In contrast, in the right panel of Figure 2, two payout dates are optimal as they exploit non-convexities in the impatience costs associated with single-date contracts (see dashed line). These non-convexities arise from sufficient changes in the growth rate of informativeness, or more generally, discrete information events (see also left and right panel of Figure 1). To tap “late” increases in informativeness, the optimal contract now makes a payment at a long-term date  $T_L$  to target (IC) and an additional short-term payment at date  $T_S$  to satisfy (PC) at lower impatience costs. The optimal choice of  $T_S$  and  $T_L$  generates a strict improvement over the single-date contract that pays out exclusively at date  $T_1(a)$  (see red circle in right panel).

### 2.2.3 Optimal contracts

Synthesizing the cases with (PC) binding and (PC) slack, we can now characterize optimal  $\mathcal{C}_{MI}$ -contracts. Together with the conditions for the optimality of  $\mathcal{C}_{MI}$ -contracts derived in Lemma 1 we thereby obtain a characterization of optimal contracts,  $B^*$  and  $w_t^*$ , based on the solution to Problem 1\*.<sup>23</sup> We define the corresponding optimal payout dates as  $T^*(a) = \{t \in [0, \bar{T}] : dw_t^* > 0\}$ .

**Theorem 1** *Suppose  $I(0|a) \leq \frac{c'(a)}{v+c(a)}$ , then (IC) is relevant for compensation costs and action  $a$  is optimally implemented with a  $\mathcal{C}_{MI}$ -contract.*

- 1) *If  $v \leq \bar{v} = \frac{c'(a)}{I(T_{RE}(a)|a)} - c(a)$ , (PC) is slack, the unique optimal payout date is  $T^*(a) = T_{RE}(a)$  as defined in (7) and the size of the compensation package is  $B^* = \frac{c'(a)}{I(T_{RE}(a)|a)}$ .*
- 2) *Otherwise, (PC) binds, so that  $B^* = v + c(a)$  and  $I_{\mathcal{C}} = \frac{c'(a)}{v+c(a)}$ . Payments are optimally made at maximally two payout dates  $T^*(a)$  as characterized in Lemma 2.*

<sup>22</sup> Appendix B.3 lists the concrete characterization of optimal payout times for several parametric specifications of the arrival time distribution in Example 2.

<sup>23</sup> For example, if a  $\mathcal{C}_{MI}$ -contract with two payment dates,  $T_S(a)$  and  $T_L(a)$ , is optimal, then  $w_t^* = 0$  for  $t < T_S$ ,  $w_t^* = \frac{I(T_L(a) - \frac{c'(a)}{v+c(a)})}{I(T_L(a) - I(T_S|a)}$  for  $T_S \leq t < T_L$  and  $w_t^* = 1$  for  $T \geq T_L$ .

Theorem 1 summarizes the intuitive characterization of the timing of optimal  $\mathcal{C}_{MI}$ -contracts in general information environments. It remains to explicitly characterize optimal contracts in the (less interesting) case when  $\mathcal{C}_{MI}$ -contracts do not apply and explain the intuition for the threshold level for  $I(0|a)$  in Theorem 1.

**Lemma 3** *If  $I(0|a) > \frac{c'(a)}{v+c(a)}$ ,  $\mathcal{C}_{MI}$ -contracts do not apply. (PC) binds and all payments are made at time 0,  $w^*(0) = 1$ , so that  $W(a) = B^* = v + c(a)$ .*

Intuitively, if the principal receives sufficiently precise signals at time 0 (and  $v > 0$ ) the (IC) constraint becomes irrelevant for compensation costs. To understand the role of the threshold value  $I(0|a) = \frac{c'(a)}{v+c(a)}$ , suppose that the principal makes the minimum size of the compensation package required by (PC),  $B = v + c(a)$ , contingent on the most informative history at time 0,  $h_{MI}^0$ . Then, the marginal benefit of increasing the action to the agent is  $I(0|a)(v + c(a))$ . If this exceeds the marginal cost,  $c'(a)$ ,  $\mathcal{C}_{MI}$ -contracts would provide excessive incentives. An optimal contract needs to “dilute” the contract’s informativeness sufficiently so as to ensure that the weighted average likelihood ratio of date-0 performance signals satisfies  $\frac{c'(a)}{v+c(a)}$ . While the timing is uniquely determined, the associated histories (signals) for which bonuses are stipulated are typically not uniquely identified.

We have now completely characterized optimal compensation contracts to implement any given action  $a$ . The associated wage cost to the principal, the solution to Problem 1, follows immediately:

$$W(a) = \begin{cases} \frac{c'(a)}{I(T_{RE}|a)} e^{\Delta_r T_{RE}} & v \leq \bar{v} \\ (v + c(a)) C\left(\frac{c'(a)}{v+c(a)} \middle| a\right) & v > \bar{v} \end{cases}.$$

#### 2.2.4 Comparative statics and discussion

Motivated by the discussion revolving around “short-termist” compensation contracts for CEOs (cf. Bebchuk and Fried (2010)) we now use our model of optimal compensation design to analyze the comparative statics of the duration of the optimal compensation package  $\int t dw_t^*$ .

**Proposition 2** *The duration of the compensation package is*

- i) decreasing in  $v$ ,*
- ii) may be increasing or decreasing in  $a$ .*

The intuition for these comparative statics all follow from the fact that the size of the compensation package,  $B$ , and more informative performance signals,  $I_{\mathcal{C}}$ , are substitutes

for providing incentives to the agent,  $BI_\varphi = c'(a)$ . When an increase in the agent's outside option exogenously raises the size of pay, this substitutability implies that the principal optimally shortens the duration of the compensation package, such as to reduce contract informativeness (strictly so if (PC) binds).

The comparative statics for effort incentives  $a$  are more subtle, mainly because informativeness itself varies with the action. Consider, first, the case when (PC) is slack such that the principal is unconstrained to adjust both  $B$  and  $I_\varphi = I(T_{RE}(a)|a)$  to provide incentives. The optimal combination and as such the optimal duration choice  $T_{RE}(a)$  then depend on whether the principal learns faster under high or low effort, which depends on the specification of the signal process. To illustrate this point, it is convenient to assume that the first-order condition in (8) applies. Then, the sign of the comparative statics of contract duration  $T_{RE}(a)$  in  $a$  depends on whether the (log) growth rate of informativeness,  $\frac{d \log I}{dt}$ , increases or decreases in  $a$ , i.e.,

$$\operatorname{sgn} \left( \frac{dT_{RE}(a)}{da} \right) = \operatorname{sgn} \left( \frac{d}{da} \frac{d \log I(t|a)}{dt} \Big|_{t=T_{RE}(a)} \right). \quad (9)$$

In Appendix B.3 we illustrate, using different parametric survival distributions within our Example 2, that all comparative statics are generically possible.<sup>24</sup> Hence, it is misguided to conclude that short-term payout times are necessarily reflective of poor incentives, i.e., low actions  $a$ . One needs to consider the entire compensation package; in particular, the size of pay. When the size of pay is determined by the agent's outside option, i.e., when (PC) binds, it follows from incentive compatibility that, in order to implement a higher action  $a$ , the timing of pay, as represented by  $w_t^*(a)$ , has to adjust such that contract informativeness increases. Still, this does not imply that the duration of the compensation package has to increase as long as the principal can adjust the timing of pay along several margins, allowing him to exploit differential (local) changes in informativeness, as is the case when two payout dates are optimal.<sup>25</sup> Thus, overall, the duration of the compensation package alone is not necessarily a good indicator for the incentives provided by the contract.

<sup>24</sup>The exponential distribution represents the knife-edge case, where the optimal payout time,  $T_{RE} = \frac{1}{\Delta_r}$ , is independent of the action, cf. Corollary B.1.

<sup>25</sup>Formally, a higher contract informativeness  $I_\varphi = \int_0^T I(t|a)dw_t$  does not necessarily imply a higher duration  $\int_0^T tdw_t$ . If a single payout date is optimal, i.e.,  $I_\varphi = I(T_1(a)|a)$ , then we can show, under some mild regularity conditions, that contract duration is an increasing function of  $a$ , i.e.,  $T_1'(a) \geq 0$  (cf. Lemma A.2 in the Appendix).

**Discussion.** Two features of our setting were essential in making our compensation design analysis particularly tractable, risk-neutrality and the validity of the first-order approach. First, we consider the validity of the first-order approach. Given the features of the optimal contract we can now state sufficient conditions for its validity

**Lemma 4** *If  $L_t(\tilde{a}|h_{MI}^T(a))$  is strictly concave in  $\tilde{a}$  for the payout dates  $T$  characterized in Theorem 1, then the first-order approach is valid for action  $a$ .*

The condition is reminiscent of the convexity of the distribution function condition (CDFC) in static models (cf. e.g., Rogerson (1985)) and, hence, subject to similar concerns about restrictiveness (cf. Renner and Schmedders (2015)). Of course, such technical issues do not arise in the case of a binary action set. Our characterization of optimal contracts in Theorem 1 extends one-to-one to this case (cf. Appendix B.1).

Second, risk-neutrality plays a key role in our model. In the absence of risk-sharing considerations, only maximally informative histories are relevant for compensation design, so that the incentive properties of any signal process can be encoded in an increasing function  $I$ . The implications of such incentives-schemes are stark, in that high rewards may be stipulated for low-probability events. Unlike in our Examples 1 and 2, the history  $h_{MI}^t(a)$  may even have zero mass so that an unbounded bonus would need to be stipulated to satisfy (IC). In such cases, it is economically sensible to impose payment bounds.<sup>26</sup> Such bounds may either be literally interpreted as capturing a resource constraint, such as the principal's limited liability, or the agent's risk-aversion in the simplest possible fashion (see e.g., Plantin and Tirole (2015)). The main intuition underlying our results is robust even in the presence of such bounds. The implication is that as the bound starts to interfere with the unconstrained bonus size, the principal will start rewarding the agent for histories associated with lower likelihood ratios for a given  $t$ , or, eventually, also choose a larger selection of payout dates. Concretely, suppose PC is slack, then there exists a cutoff value  $\kappa_{IC}$  such that the principal rewards the agent at  $t$  following history  $h^t$  if and only if  $\tilde{I}_t(h^t|a) \geq \frac{1}{\kappa_{IC}}e^{\Delta_r t}$  where  $\tilde{I}_t(h^t|a) = \frac{d \log L_t(a|h^t)}{da}$  refers to the date- $t$  likelihood ratio of history  $h^t$ .<sup>27</sup> While the payout time need no longer be unique, the main underlying intuition is robust. The principal's rent-extraction motive (measured in likelihood ratio units) is balanced against the costs resulting from the agent's liquidity needs.

Given the robust intuition behind optimal contract design, we will now analyze the optimal action choice that the principal induces and analyze the effects of mandatory

<sup>26</sup> Then, the solution to Problem 1\* would represent a lower bound on compensation costs.

<sup>27</sup> If the pay cap does not constrain the principal, then  $\kappa_{IC} = \frac{e^{\Delta_r T_{RE}}}{I(T_{RE}|a)}$ . Otherwise,  $\kappa_{IC} > \frac{I(T_{RE}|a)}{e^{\Delta_r T_{RE}}}$ .

deferral regulation.

### 3 Action choice and the effects of deferral regulation

#### 3.1 Unconstrained action and rationale for regulation

So far, the entire analysis has focused on the principal’s costs to induce a given action,  $W(a)$ . To determine the equilibrium action  $a^*$  we also need to account for the benefits of an action to the principal, which we capture by a strictly increasing and concave bounded function  $\pi(a)$ . Here,  $\pi(a)$  may correspond to the *present value* of the (gross) profit streams under action  $a$ , but could also be interpreted more broadly as the principal’s utility derived from action  $a$ . Then, absent other constraints, the equilibrium action maximizes  $\pi(a) - W(a)$ , and, given a solution  $a^*$ , which we assume to be interior, the chosen payout times are characterized by Theorem 1 (and Lemma 3).

#### Problem 2 (Equilibrium action)

$$a^* = \arg \max_{a \in \mathcal{A}} \pi(a) - W(a). \quad (10)$$

As discussed in the Introduction, there are various reasons for why the principal optimal equilibrium action  $a^*$  may not be socially optimal, such that there is scope for regulatory intervention. For instance, we now suppose that total welfare comprises an additional term  $-x(a)$ , not accounted for by the principal,<sup>28</sup> where  $x(a) > 0$  is strictly decreasing in  $a$ , so that  $a^*$  is too low from society’s perspective. Motivated by recent regulatory initiatives in the financial sector, we then suppose that the regulator wants to correct, i.e., increase, the equilibrium action through imposing a minimum deferral period. More formally, regulation now imposes the following, additional constraint on Problem 1

$$b_t = 0 \quad \forall t < \tau. \quad (\text{R})$$

While we are a priori agnostic about whether such deferral regulation is optimal or not (see discussion in Conclusion), this specific tool has been rationalized by regulators on

---

<sup>28</sup>For example, if we interpret the principal as the board of a bank,  $x(a)$  may literally represent an externality of bank failure on other banks (see, e.g., Allen and Gale (2000)), it may stem from bank shareholders exploiting the regulatory guarantees implied by deposit insurance (see, e.g., Merton (1977)), or originate from a shorter time horizon of the bank’s board than the one of society (see, e.g., Guttentag and Herring (1984)). Finally, even if bank shareholders are aligned with society, it is possible that the principal, the board, does not maximize shareholder value due to a corporate governance problem (see, e.g., Kuhnen and Zwiebel (2009)).

the grounds that existing compensation packages are “too short-termist” and “only time will tell” whether bankers act prudently.<sup>29</sup>

To ensure that the fundamental rationale behind deferral regulation is sound, we assume that the regulators’ diagnosis of “too short-termist” compensation packages is correct, i.e., we restrict attention to information environments where the unconstrained optimal duration of pay is indeed strictly increasing in the action. While Proposition 2 shows that this assumption constrains the set of possible information environments, it rules out a trivial way in which minimum deferral regulation could “backfire.” Suppose payout times were decreasing in the action, then a minimum deferral period  $\tau$  would perversely constrain the principal in implementing high action levels, but not for low action levels. In such an information environment there is, thus, no scope for the tool of *minimum* deferral regulation.

Our primary goal is now a comparative statics analysis of the action choice in response to changes in the minimum deferral period  $\tau$ , rather than the broader goal of optimal regulation. To streamline this analysis, we make the following additional technical assumptions:<sup>30</sup>

**Assumption 1** *Technical assumptions on information environment:*

- 1)  $I(t|a)$  is differentiable in both arguments on  $[0, \infty) \times \mathcal{A}$  and, for any  $a \in \mathcal{A}$ , a strictly increasing function of time with  $I(0|a) = 0$ .
- 2) For any  $a \in \mathcal{A}$ ,  $C(I_{\mathcal{E}}|a)$  is strictly convex in  $I_{\mathcal{E}}$ .

Condition 1) is sufficient to ensure that the principal uses  $\mathcal{C}_{MI}$ -contracts in the absence of regulation (since  $I(0|a) = 0$ , cf. Theorem 1), that deferral always leads to strictly more informativeness and that we can consider the limit  $\tau = \infty$ . Condition 2) implies that there is always a unique and bounded optimal payout time in the absence of regulation,  $T^*(a)$ . Moreover, if regulation constrains the principal to implement a certain action  $a$ , he responds by choosing a payout time that is the earliest feasible deferral period. It thus rules out the surprising (but presumably less relevant) case that the principal chooses a contract duration strictly greater than  $\tau$  under a minimum deferral requirement of  $\tau$  years, even though he would have chosen a duration of  $T^*(a) < \tau$  in the absence of

---

<sup>29</sup> Cf., e.g., the following quote by Mark Carney, Governor of the Bank of England from November 17, 2014: “Compensation schemes overvalued the present and heavily discounted the future, encouraging (...) short-termism. (...) In the UK, we have introduced a remuneration code prescribing that payment of bonuses must be deferred for a minimum of three years and (...) be exposed to clawback for up to seven years.”

<sup>30</sup> In an earlier working paper we did not make these assumptions and obtained the same general insights. The analysis is available from the authors upon request.

regulation (see Example B.4 and Figure 7 in Appendix B.3).<sup>31</sup>

## 3.2 Comparative statics analysis

For ease of exposition, we initially consider the effects of deferral regulation for the case when the agent's outside option is zero so that the principal's problem is a rent-extraction problem (see Section 3.2.1). We then consider the general case allowing for  $v > 0$  (see Section 3.2.2). For each case, we first analyze how the principal optimally restructures compensation contracts constrained by regulation in order to implement a given action  $a$  and then analyze the effect on the equilibrium action choice.

### 3.2.1 Pure rent-extraction case

When the agent's outside option is 0, the optimal payout time in the absence of regulation is given by  $T^*(a) = T_{RE}(a)$  (see Theorem 1). Thus, regulation constrains the principal's optimal choice of payout times for a given action  $a$  as soon as  $\tau > T^*(a)$ . Let  $T^*(a|\tau) := \arg \min_{t \geq \tau} \frac{e^{\Delta r t}}{I(t|a)}$  denote the optimal payout time given deferral period  $\tau$ , then the strict-convexity condition of Assumption 1 implies that

$$T^*(a|\tau) = \max \{T^*(a), \tau\}. \quad (11)$$

For ease of notation, if the conditioning variable  $\tau$  is omitted, we refer to the case without regulation, e.g.,  $T^*(a|0) = T^*(a)$ .

**Lemma 5** *If  $v = 0$  and  $\tau > T^*(a)$  the optimal contract to implement action  $a$  is a  $\mathcal{C}_{MI}$ -contract with payout time  $\tau$  and a reduced size of the compensation package satisfying*

$$B(\tau) = \frac{c'(a)}{I(\tau|a)} < B(0) = \frac{c'(a)}{I(T^*(a)|a)}. \quad (12)$$

Thus, as soon as the regulatory constraint interferes with the principal's unconstrained choice, the principal uses the increase in informativeness induced by forced deferral above and beyond  $T^*(a)$ ,  $I(\tau|a) > I(T^*(a)|a)$ , to reduce the size of the agent's compensation package  $B$ . By optimality of the original timing choice  $T^*$ , however, this reduction in  $B$  must be more than outweighed by the increased impatience cost. Compensation costs to implement a given  $a$  increase,  $W(a|\tau) > W(a)$ . Since the unconstrained optimal payout

---

<sup>31</sup>This result is reminiscent of Jewitt, Kadan, and Swinkels (2008) who show that minimum wages may harm the principal even when the agent receives more than the minimum wage in the (constrained) optimal contract.

times are (by assumption) strictly increasing in the action, there exists a cutoff action  $\underline{a}(\tau)$  such that regulation only constrains the principal's choice of compensation contract for  $a < \underline{a}(\tau)$ . We can then completely characterize the constrained wage cost function

$$W(a|\tau) = \begin{cases} e^{\Delta_r \tau} \frac{c'(a)}{I(\tau|a)} & a < \underline{a}(\tau) \\ W(a) & a \geq \underline{a}(\tau) \end{cases}. \quad (13)$$

Facing this constrained optimal compensation cost function  $W(a|\tau)$  the principal now chooses the action that maximizes his net profits

$$a^*(\tau) = \arg \max_{a \in \mathcal{A}} \pi(a) - W(a|\tau). \quad (14)$$

Note that the considered regulation only affects the cost side of principal's preferences,  $W(a|\tau)$ , but does not affect the benefit side,  $\pi(a)$ , i.e., it does not directly alleviate the distortion in the principal's preferences,  $x(a)$ . A trivial implication is that our main objective, the comparative statics analysis of the equilibrium action  $a^*(\tau)$  in  $\tau$ , is robust to the source and specification of  $x(a)$ . For ease of exposition, we assume that  $\pi(a) - W(a|\tau)$  is a differentiable and strictly quasiconcave function of  $a$  such that  $a^*(\tau)$  is uniquely determined from (14).<sup>32</sup> Hence, the comparative statics of  $a^*$  in  $\tau$  solely depend on whether deferral regulation raises or lowers *marginal* compensation costs  $\frac{\partial W(a|\tau)}{\partial a}$ .

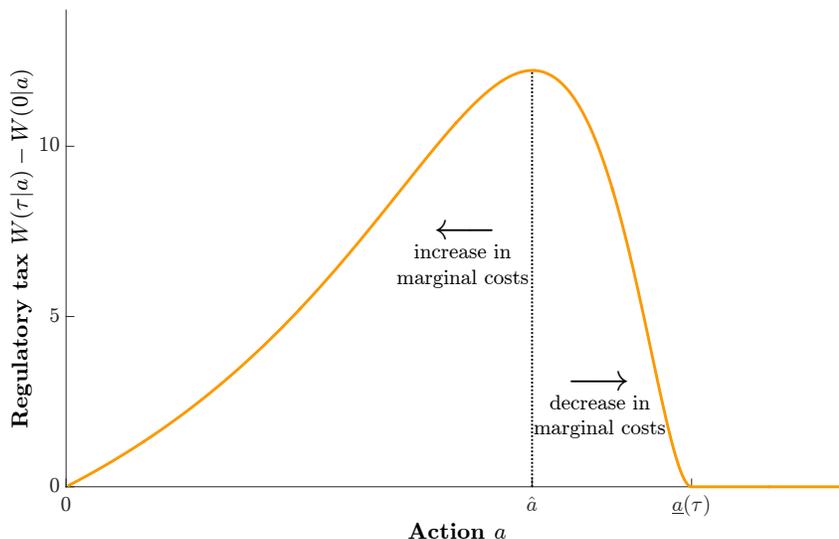
To understand the effect of deferral regulation on marginal compensation costs, it is useful to first analyze the properties of the regulatory tax,  $\Delta W(a|\tau) = W(a|\tau) - W(a|0)$ , which measures the increase in the compensation costs, at a given  $a$ , induced by a given deferral period  $\tau$ . If the regulatory tax *increases* (decreases) in the action  $a$ ,  $\frac{\partial \Delta W(a|\tau)}{\partial a} > 0$ , then marginal compensation costs under deferral regulation are higher compared to the absence of regulation, i.e.,  $\frac{\partial W(a|\tau)}{\partial a} > \frac{\partial W(a|0)}{\partial a}$ .

**Lemma 6** *Suppose  $T_{RE}(0) < \tau < T_{RE}(\bar{a})$ . Then, there exists  $\underline{a}(\tau) \in (0, \bar{a})$  such that the regulatory tax  $\Delta W(a|\tau)$  is equal to zero for  $a = 0$  and  $a \geq \underline{a}(\tau)$ , while it is strictly positive for  $a \in (0, \underline{a}(\tau))$ . For  $a \in (0, \underline{a}(\tau))$ ,  $\Delta W(a|\tau)$  is strictly increasing in  $a$  for  $a$  sufficiently small and strictly decreasing for  $a$  sufficiently close to  $\underline{a}(\tau)$ .*

The intuition behind the non-monotonicity of the regulatory tax (see Figure 3) results from two robust effects that jointly determine how deferral regulation operates. First, since we restrict attention to information environments where the unconstrained optimal

---

<sup>32</sup>See the example in the Introduction for robustness of intuition with a discrete action set and no differentiability assumptions. The advantage of the continuous action set employed in the main part of the paper is that actions react to *marginal* changes in the deferral period.



**Figure 3. Regulatory tax:** The figure plots the regulatory tax,  $\Delta W(a|\tau) = W(a|\tau) - W(a|0)$ , as a function of  $a$  for a given level of  $\tau > T^*(0)$ , for Example 2 with a lognormal arrival time distribution as specified in Example B.3 in Appendix B.3. The chosen parameter values are  $\Delta_r = 1/3$ ,  $\sigma = 1$  and  $\tau = 6.5$  with quadratic costs  $c(a) = \frac{1}{2}a^2$ .

payout time  $T^*(a)$  is increasing in the action, the deviation from the optimal timing of pay forced by regulation,  $\tau - T^*(a)$ , is decreasing in the action. In particular, there exists a unique threshold value  $\underline{a}(\tau)$  so that minimum deferral regulation does not constrain the principal for  $a \geq \underline{a}(\tau)$ , i.e.,  $\Delta W(a|\tau) = 0$  for high action levels. Second, the size of the tax burden interacts with the agency problem in that it is *proportional* to the size of pay required to incentivize an action. Since the size of the compensation package increases with the induced action, so does ceteris paribus the tax burden. In particular, when no incentives need to be provided, i.e.,  $a = 0$ , a forced deferral is not costly to the principal.<sup>33</sup> Taken together, in the intermediate range between 0 and  $\underline{a}(\tau)$ , regulation constrains the principal, resulting in  $\Delta W(a|\tau) > 0$ . By continuity, we thus obtain the formal results stated in Lemma 6, which is the foundation for the following Proposition.

**Proposition 3** *Consider the case with  $v = 0$ . Then, a sufficiently small increase in  $\tau$  above the unconstrained optimal payout time  $T^*(a^*(0))$  induces an increase in the action, i.e.,  $a^*(\tau) > a^*(0)$ . However, for  $\tau$  sufficiently high, the action satisfies  $a^*(\tau) < a^*(0)$  with  $\lim_{\tau \rightarrow \infty} a^*(\tau) = 0$ .*

If the minimum deferral period is sufficiently close to the principal's unconstrained timing choice  $T^*(a^*(0))$ , then  $a^*(0)$  is sufficiently close to  $\underline{a}(\tau)$ . Lemma 6 then im-

<sup>33</sup>Formally, as  $c'(0) = 0$ , (13) implies that  $W(0|\tau) = 0 = W(0)$ .

plies that the regulatory tax is decreasing in  $a$  for  $a \in (a^*(0), \underline{a}(\tau))$ , or equivalently, that marginal compensation costs are lower compared to the absence of regulation, i.e.,  $\frac{\partial W(a|\tau)}{\partial a} < \frac{\partial W(a|0)}{\partial a}$ . Thus, for the given deferral period  $\tau$  it must be strictly beneficial for the principal to increase the action above and beyond the original unconstrained action choice as the marginal benefit exceeds the marginal costs evaluated at  $a^*(0)$

$$\pi'(a^*(0)) = \left. \frac{\partial W(a|0)}{\partial a} \right|_{a=a^*(0)} > \left. \frac{\partial W(a|\tau)}{\partial a} \right|_{a=a^*(0)}.$$

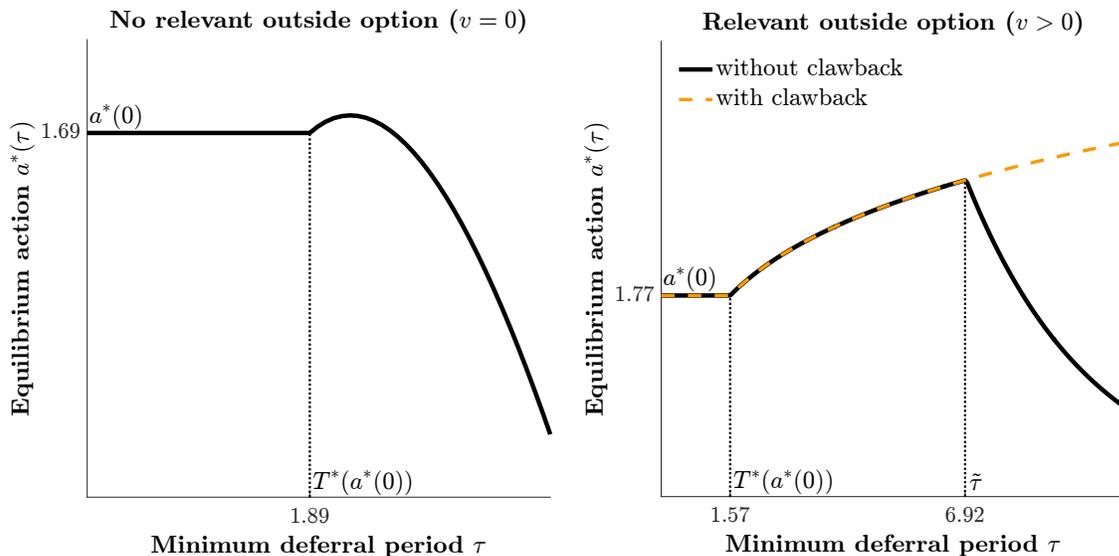
This explains the first result. In the special case plotted in Figure 3, the regulatory tax is single-peaked and, hence, strictly decreasing in  $a$  for all  $a > \hat{a}$ . That is, as long as the unconstrained action satisfies  $a^*(0) > \hat{a}$ , regulation works, i.e.,  $a^*(\tau) > a^*(0)$ .

In contrast, sufficiently high deferral periods always backfire (marginal compensation costs are increased compared to the absence of regulation). In the context of Figure 3 the given deferral period  $\tau$  backfires if  $a^*(0) < \hat{a}$  (by the reverse argument of before). Finally, in the “trivial” limit, as  $\tau \rightarrow \infty$ , the *marginal* compensation cost of inducing *any* action that requires some incentive pay,  $a > 0$ , approaches infinity. Thus, the action must go to zero.

Summing up, for the case with  $v = 0$ , imposing a binding mandatory deferral period indeed leads to a strictly higher level of  $a$  in equilibrium as long as the imposed deferral period is not too onerous, and will “backfire” otherwise. The left panel of Figure 4 plots an example specification illustrating these results. We will show next that the qualitative effects of deferral regulation on the equilibrium action are robust, i.e., extend to the case of  $v > 0$ , as depicted in the right panel of Figure 4 (albeit with an additional qualification regarding the use of clawback clauses, which we address below).

### 3.2.2 General case

We now allow for a positive outside option  $v > 0$ . First, we again consider the compensation design problem under regulation, i.e., (constrained) optimal contracts to implement a fixed action  $a$ . Clearly, as long as the size of the compensation package implied by the solution to the pure rent extraction problem,  $B(\tau) = \frac{c'(a)}{I(\tau|a)}$ , exceeds  $v + c(a)$ , (PC) is still slack and Lemma 5 also characterizes optimal contracts in the presence of a strictly positive outside option. That is, in response to regulation, the principal continues to use  $\mathcal{C}_{MI}$ -contracts but adjusts the size of pay. The interesting (and new) case to study is the design of constrained optimal contracts when the agent’s participation constraint binds. Then, in the absence of regulation, Lemma 2 and Assumption 1 imply that the principal would choose a  $\mathcal{C}_{MI}$ -contract with payout time  $T_1$  such that  $I(T_1(a)|a) = \frac{c'(a)}{v+c(a)}$ .



**Figure 4. Equilibrium action:** The figure plots the equilibrium action as a function of the minimum deferral period  $\tau$  for the case of  $v = 0$  (left panel) and  $v > 0$  (right panel). The information environment is as in Example 2 with a lognormal arrival time distribution (cf. Example B.3 in Appendix B.3). The chosen parameter values are  $\Delta_r = 0.7$ ,  $\sigma = 1$ , with costs  $c(a) = \frac{1}{30}a^3$  and gross profits  $\pi(a) = 10a$ , and, for the right panel,  $v = 1.5$ .

The following Lemma shows how the principal optimally restructures this compensation contract when facing (binding) deferral regulation.

**Lemma 7** *Suppose that regulation is sufficiently stringent, i.e.,  $I(\tau|a) > \frac{c'(a)}{v+c(a)}$ . Then (PC) binds and regulation prohibits the principal from implementing action  $a$  with a  $\mathcal{C}_{MI}$ -contract. All payouts are made at time  $\tau$  implying that  $W(a|\tau) = e^{\Delta_r \tau} (v + c(a))$ .*

Intuitively, when  $I(\tau|a) > \frac{c'(a)}{v+c(a)}$ , making the entire pay required by (PC),  $v + c(a)$ , contingent on the most informative history  $h_{MI}^\tau$  provides excessive incentives, i.e., violates (IC). Hence, in order to implement such an action when the size of pay is fixed by the agent's outside option, the principal must adjust the contingency of pay and pay out part of the compensation following less informative histories (cf. Lemma 3). Within the context of Example 2, this implies that the principal must stipulate some payouts to the agent for histories with a failure. Moreover, since (IC) then is irrelevant for compensation costs, all payments are made as early as possible to minimize impatience costs, i.e., at date  $\tau$ .

While we have so far considered a given action  $a$ , we next analyze whether the condition of Lemma 7 can also apply for the action that the principal optimally implements in equilibrium. Building on the intuition with slack (PC), one may conjecture that in-

ducing such an action cannot be optimal since not all pay is used to provide incentives, and that the principal, hence, could profitably implement a higher  $a$  by switching to a  $\mathcal{C}_{MI}$ -contract. However, Proposition 4 shows that these conjectures are wrong when the agent has a relevant outside option,  $v > 0$ .

**Proposition 4** *The results of Proposition 3 on the effect of mandatory deferral apply irrespective of whether  $v = 0$  or  $v > 0$ . In addition, when  $v > 0$  there exists a threshold  $\tilde{\tau}$  such that the principal no longer uses  $\mathcal{C}_{MI}$ -contracts for  $\tau > \tilde{\tau}$ . For  $\tau > \tilde{\tau}$ ,  $a^*(\tau)$  is strictly decreasing in  $\tau$ .*

The general mechanism for why regulation initially works (and eventually backfires) is very similar to the pure rent extraction case. To focus on the new case, we suppose (PC) binds in the absence of regulation so that the single optimal payout date is given by  $T_1(a^*(0))$ . Since  $T_1(a)$  is strictly increasing in  $a$ , a minimum deferral period of  $\tau$  will again “tax” low actions, i.e., increase compensation costs for actions  $a < \underline{a}(\tau)$ , while not constraining optimal contracts for sufficiently high actions  $a \geq \underline{a}(\tau)$ .<sup>34</sup> Thus, if the binding deferral period  $\tau$  is sufficiently close to  $T^*(a^*(0))$ , the principal will optimally implement a higher action,  $a^*(\tau) = \underline{a}(\tau)$ , which is strictly increasing in  $\tau$ . Deferral regulation again “works” as a tax on low actions (see right panel Figure 4 for  $T_1(a^*(0)) \leq \tau \leq \tilde{\tau}$ ). The principal fully uses the increase in informativeness imposed by regulation to provide incentives and implements the lowest possible action that can be implemented with a  $\mathcal{C}_{MI}$ -contract,  $a^*(\tau) = \underline{a}(\tau)$ .<sup>35</sup>

However, deferral regulation must eventually backfire since the induced increase in the action becomes increasingly costly to the principal. First, (PC) implies that an increase in  $a$  requires a larger compensation package,  $v + c(a)$ , since the agent has to be compensated for higher effort costs. Second, further deferral makes the entire compensation package (including  $c(a)$ ) increasingly costly to the principal due to relative impatience. Thus, as regulation becomes sufficiently onerous,  $\tau > \tilde{\tau}$ , the principal eventually implements actions  $a^*(\tau) < \underline{a}(\tau)$  to reduce the size of the tax burden induced by deferral regulation. However, to implement such actions he needs to deviate from  $\mathcal{C}_{MI}$ -contracts (see Lemma 7) implying wage costs of  $W(a|\tau) = e^{\Delta r \tau} (v + c(a))$ . For  $\tau > \tilde{\tau}$ , the optimal action is thus fully characterized by the first-order condition  $\pi'(a) = e^{\Delta r \tau} c'(a)$  so that  $a^*(\tau)$  is strictly decreasing in  $\tau$  (see right panel of Figure 4 for  $\tau > \tilde{\tau}$ ).

<sup>34</sup> Here,  $\underline{a}(\tau)$  is implicitly defined by  $I(\tau|a) = \frac{c'(a)}{v+c(a)}$ , such that  $a^*(0) = \underline{a}(T_1(a^*(0)))$ . The proof of Proposition 4 shows that  $\underline{a}(\tau)$  is strictly increasing in  $\tau$ .

<sup>35</sup> Hence, albeit regulation has thus, for sufficiently small  $\tau$ , positive effects regardless of whether (PC) binds or not, there is a subtle difference. When (PC) does not bind, the new (higher) value  $a^*(\tau)$  is obtained from the principal’s first-order condition, while otherwise it is pinned down by binding (PC) and (IC).

Thus, the general effects of deferral regulation on the action choice are robust regardless of whether (PC) binds or not. However, both cases crucially differ in terms of the contract adjustments that the principal undertakes in response to deferral regulation. Only when (PC) binds, does the principal deviate from  $\mathcal{C}_{MI}$ -contracts. Arguably, granting the principal such flexibility to circumvent deferral regulation is not in the policymakers' interest. That is, if the regulator chooses a stringent deferral period  $\tau > \tilde{\tau}$  with the goal of increasing the action, then regulation should also constrain the contingency of pay. The following Corollary confirms this intuition.

**Corollary 1** *Suppose that  $v > 0$  and a sufficiently long deferral period  $\tau > \tilde{\tau}$  is implemented. Then, the induced equilibrium action is strictly higher when regulation also requires the principal to use  $\mathcal{C}_{MI}$ -contracts.*

In our Example 2, where survival is the most informative history, such an additional regulation is straightforward to implement by forcing the principal to pay out compensation only when no previous failure occurred.<sup>36</sup> Thus, with this additional restriction any promised compensation must then be forfeited in case of failure, which is akin to a mandatory malus clause that is observed in practice (cf. the examples of financial regulation in the Introduction). A malus clause is the technical definition of a clawback of promised, but yet unpaid bonus payments, i.e., from an escrow account.<sup>37</sup> The right panel of Figure 4 illustrates the effects of such a clause for  $\tau > \tilde{\tau}$  within Example 2.<sup>38</sup> This completes our formal analysis of deferral regulation.

Before concluding, it is useful to come back to our discussion in the literature review and to compare minimum deferral regulation with a regulation that forces the principal to invest in a better information system in the sense of Gjesdal (1982) or Kim (1995), say an (improved) IT system, at some lump-sum cost  $k$ . Similar to such a forced lump-sum investment, mandatory deferral indeed implies that the information system at the time of payout has improved. However, both regulations fundamentally differ in terms of their associated costs and, hence, their effect on the equilibrium action. For the case of an investment in the IT system, the respective costs are lump-sum and would, thus, arise *independently* of the actual incentives that the principal wishes to provide to the agent.

---

<sup>36</sup>This restriction must then clearly extend beyond  $\tau$ , so as not be circumvented again by mitigating incentives through rewarding failure.

<sup>37</sup>In practice, most “clawbacks” refer to maluses, since the retrieval of paid out bonuses is difficult for the same reason as it is impossible in the model. The agent has already consumed it.

<sup>38</sup>The combination of high  $\tau$  and a restriction to  $\mathcal{C}_{MI}$ -contracts seems to allow policymakers to induce arbitrarily high values of  $a$ , possibly leading to losses on the side of the principal. Since, realistically, the principal also has a participation constraint he would no longer wish to employ the agent for sufficiently high values of  $\tau$ .

Instead, when additional information arises from longer deferral, the respective costs are endogenously linked to these incentives as the costs arise from the agent’s relative impatience and therefore increase with the size of the compensation package  $B$  that is needed to provide these incentives. Ultimately, the principal can reduce the costs of deferral regulation by paying the agent less, which is the very reason why pure deferral regulation eventually backfires.

## 4 Conclusion

Our paper aims to shed light on the optimal duration of pay in principal-agent settings where performance signals arrive over time and the associated increase in informativeness has to be balanced against the agent’s liquidity needs. By allowing for general information systems in a tractable way, our theory contributes towards a better understanding of the general implications of such “only time will tell” environments for optimal contracting. We then use this framework to analyze the effects of recent regulatory interventions in the financial sector, such as minimum deferral periods for bonuses, and show that moderate interventions can effectively nudge shareholders to incentivize more long-run actions, while sufficiently long minimum deferral periods unambiguously backfire.

The broader take-away of our paper is that the observed duration of pay may reflect an optimal trade-off rather than necessarily a corporate governance failure. Our results suggest that the empirical analysis of compensation contracts ought to relate the timing dimension of pay also to variation of the nature of information arrival across industries (see e.g., Gopalan, Milbourn, Song, and Thakor (2014) for an empirical analysis of the duration of executive compensation). For example, it may be hypothesized that firms in R&D intensive industries (with year-long lags and few products) ought to design longer-duration contracts for their CEOs.

From a theoretical perspective, we see several fruitful avenues for future research. First, our focus to obtain a tractable model of the *timing* of pay, did not aim to produce realistic contracts in terms of the *contingency* of pay. This can be fixed by incorporating monotonicity constraints on the principal’s payoffs, so the optimal contract takes on the form of debt (cf. Innes (1990)), the empirically most relevant source of external financing for firms.<sup>39</sup> Our framework could then be used to derive a theory of optimal debt maturity resulting from the trade-off between a firm’s liquidity needs and the increased informativeness associated with new performance signals available to investors. In particular, it would be interesting to analyze if our moral-hazard approach produces different

---

<sup>39</sup> See Hébert (2015) for the optimality of debt in more general theoretical settings.

testable predictions than the adverse selection approach by Diamond (1991).

Second, in many real-life settings, the agent has to perform multiple actions in sequence and this will typically have repercussions on optimal contract design. We have already outlined a special case in our Appendix, where our methods can be applied one-to-one. There, the agent sequentially performs a short-run (acquisition) task and, conditional on success in the acquisition round, a long-run (effort) task. It might be worthwhile to study whether it is possible to maintain the tractability of our framework in more general repeated-action setups.

Third, our paper neither tackles the broader issue of optimal financial regulation nor does it study the industry-wide implications. It may be both worthwhile to study the interplay of deferral regulation with other interventions in the structure of compensation design, such as pay caps (e.g., Thanassoulis (2012) and Bénabou and Tirole (2016)), as well as with interventions that directly target shareholders' skin in the game, such as capital requirements. Under which circumstances is deferral regulation part of the optimal policy mix? What are the effects of such regulation on the outside option of employees in the financial sector? We leave these important questions for future work.

## Appendix A Proofs

**Proof of Lemma 1.** We will start with some useful definitions: Let  $B$  denote the agent's expected time-0 valuation of the compensation package, i.e.,

$$B := \mathbb{E} \left[ \int_0^{\bar{T}} e^{-r_A t} db_t \middle| a \right],$$

and denote by  $w_t$  the fraction of the compensation package that the agent derives from expected payouts up to time  $t \leq \bar{T}$ , i.e.,

$$w_t := \mathbb{E} \left[ \int_0^t e^{-r_A s} db_s \middle| a \right] / B,$$

so that  $w_{\bar{T}} = \int_0^{\bar{T}} dw_t = 1$ . Further, define for any  $t$  and  $h^t$  the likelihood ratio as

$$\tilde{I}_t(h^t|a) = \frac{d \log L_t(a|h^t)}{da}, \tag{15}$$

and denote, for each given  $t$ , the maximal likelihood ratio by  $I(t|a) := \max_{h^t \in H^t} \tilde{I}_t(h^t|a)$  which exists by assumption. Note that the incentive constraint can then be written as

$$B \left( \int_0^{\bar{T}} \left[ \int_{H^t} \tilde{I}_t(h^t|a) d\gamma_t(h^t) \right] dw_t \right) = c'(a), \quad (16)$$

for some  $\gamma_t(h^t)$  satisfying  $d\gamma_t(h^t) \geq 0$  and  $\int_{H^t} d\gamma_t(h^t) = 1$ .

The following Lemma shows when (IC) is relevant for compensation costs.

**Lemma A.1** *The shadow value on the incentive constraint,  $\kappa_{IC}$ , is zero if and only if  $I(0|a) \geq \frac{c'(a)}{v+c(a)}$ .*

**Proof of Lemma A.1** From (PC) and (LL) together with differential discounting we have that  $W(a) \geq v + c(a)$ . Hence, we need to show that  $W(a) = v + c(a)$  if and only if  $I(0|a) \geq \frac{c'(a)}{v+c(a)}$ . To show sufficiency, consider the contract delivering fraction  $\frac{c'(a)}{I(0|a)(v+c(a))}$  of total expected pay  $B = (v + c(a))$  with a bonus payment at  $t = 0$  that conditions on the realization of  $h_{MI}^0$ , and the remaining fraction  $\left(1 - \frac{c'(a)}{I(0|a)(v+c(a))}\right)$  with an unconditional payment also at  $t = 0$ .<sup>40</sup> From  $I(0|a) \geq \frac{c'(a)}{v+c(a)}$  both payments are non-negative and they clearly satisfy (PC) and (IC), while resulting in expected wage costs of  $v + c(a)$ . To show the necessary part, observe that any contract with  $W(a) = v + c(a)$  cannot feature any delay due to differential discounting, i.e., must satisfy  $w_0 = 1$ . Note further, that the contract that provides strongest incentives with date-0 payments only is, from (16), the one that makes the entire expected pay  $B$  contingent on  $h_{MI}^0$ . However, whenever  $I(0|a) < \frac{c'(a)}{v+c(a)}$ , such a contract requires  $B > v + c(a)$  in order to satisfy (IC), implying  $W(a) > v + c(a)$ . ■

Take now the case where  $\kappa_{IC} > 0$ , implying  $W(a) = B \int_0^{\bar{T}} e^{\Delta r t} dw_t > v + c(a)$ . The proof is then by contradiction. So assume that under the optimal contract there exists some  $t$  for which  $\int_{H^t \setminus h_{MI}^t} d\gamma_t(h^t) \neq 0$ . Then, there exists another feasible contract with  $\int_{H^t \setminus h_{MI}^t} d\gamma_t(h^t) = 0$  for all  $t$  and strictly lower compensation costs. To see this, observe that this new contract maximizes, for given  $w_t$  and  $B$ , the left-hand side in (16). So, assume, first, that (PC) is slack. Then, holding  $w_t$  constant, the new contract allows for a strictly lower  $B$ , thus reducing  $W(a)$ . Second, assume that (PC) binds, which, from  $\kappa_{IC} > 0$  implies that  $w_0 < 1$ . Then, holding  $B$  constant, the new contract allows to reduce some  $w_t$ ,  $t > 0$ , and increase  $w_0$  resulting in lower  $W(a)$ . **Q.E.D.**

<sup>40</sup> Put differently, this contract stipulates two payments at  $t = 0$ : A bonus  $b$  which is paid if and only if  $h_{MI}^0$  realizes and solves  $\mathbb{E}[b|a] = \frac{c'(a)}{I(0|a)}$ , as well as an unconditional payment of  $(v + c(a)) - \frac{c'(a)}{I(0|a)}$ .

**Proof of Proposition 1.** The result follows from the fact that  $\tilde{I}_t(h^t|a)$  as defined in (15) is a martingale. To see this denote by  $h^{t-}$  the history at  $t$  before the time- $t$  signal  $x_t$  has realized, i.e.,  $h^{t-} = \lim_{s \uparrow t} h^s$ . Further, define the likelihood ratio (score) for any possible time- $t$  signal  $x_t \in X_t$  conditional on history  $h^{t-}$  as  $l_t(x_t|a, h^{t-}) := \frac{d \log L_t(a|h^{t-}, x_t)}{da}$ . Then for any  $t > 0$  we have

$$\begin{aligned} \mathbb{E} \left[ \tilde{I}_t(h^t|a) \middle| h^{t-}, a \right] &= \tilde{I}_{t-}(h^{t-}|a) + \mathbb{E}_x \left[ l_t(x_t|a, h^{t-}) \middle| h^{t-}, a \right] \\ &= \tilde{I}_{t-}(h^{t-}|a), \end{aligned}$$

where we have used that the expectation of the score,  $l_t(x_t|a, h^{t-})$ , is zero (see e.g. Casella and Berger (2002)). **Q.E.D.**

**Proof of Lemma 2.** See main text. **Q.E.D.**

**Proof of Theorem 1.** It follows from Lemma A.1 that  $\kappa_{IC} > 0$  if and only if  $I(0|a) < \frac{c'(a)}{v+c(a)}$ , such that, from Lemma 1, the optimal contract must be a  $\mathcal{C}_{MI}$ -contract and hence solves Problem 1\*. Similarly, for the case where  $I(0|a) = \frac{c'(a)}{v+c(a)}$ , it follows from the arguments in the proof of Lemma A.1 that  $\kappa_{IC} = 0$  and  $W(a) = v + c(a)$  if and only if  $a$  is implemented with a  $\mathcal{C}_{MI}$ -contract, such that the optimal contract solves again Problem 1\*.<sup>41</sup>

Consider now, first, the relaxed problem ignoring (PC). Then, as shown in the main text, the optimal payout time is given by  $T_{RE}(a)$  as characterized in (7) which implies from (IC) that  $B = c'(a)/I(T_{RE}(a)|a)$ . Then (PC) is indeed satisfied, such that the solution to the relaxed problem solves the full problem, if and only if  $B \geq (v + c(a))$  which is equivalent to  $v \leq \bar{v} := c'(a)/I(T_{RE}(a)|a) - c(a)$ . Else, it is easy to show that (PC) must bind under the optimal contract, i.e.,  $B = v + c(a)$  and the optimal payout times are as characterized in Lemma 2. **Q.E.D.**

**Proof of Lemma 3.** It follows from Lemma A.1 that under the optimal contract (PC) binds and  $W(a) = v + c(a)$ . It remains to show that  $a$  is not implementable at same costs with a  $\mathcal{C}_{MI}$ -contract. Due to differential discounting it is sufficient to show that  $a$  cannot be implemented with the  $\mathcal{C}_{MI}$ -contract with  $w_0 = 1$ . To see this, note that for this contract (IC) requires that  $B = \frac{c'(a)}{I(0|a)} < v + c(a)$  violating (PC).<sup>42</sup> **Q.E.D.**

<sup>41</sup> Note that, the solution to Problem 1\* solves Problem 1 whenever a solution  $db_t$  to  $\mathbb{E}[db_t|a]e^{-rA^t} = Bdw_t$  exists for all  $t$ , given  $B$ ,  $w_t$  and  $h_{MI}^t(a)$  as implied by Theorem 1.

<sup>42</sup> The same argument implies that, due to monotonicity of  $I(t|a)$  (cf. Proposition 1), action  $a$  is not implementable with any  $\mathcal{C}_{MI}$ -contract.

**Proof of Proposition 2.** The results hold trivially when  $I(0|a) > \frac{c'(a)}{v+c(a)}$ , as then, from Lemma 3, the duration is equal to zero. We will, thus, focus on the remaining case, for which optimal contracts are as characterized in Theorem 1. Consider, first, the comparative statics in  $v$ . Clearly, when (PC) is slack, the duration  $T_{RE}(a)$  does not depend on  $v$ . When (PC) binds, the result follows directly from  $I_{\mathcal{E}} = \frac{c'(a)}{v+c(a)}$ , which is decreasing in  $v$ , together with Lemma 2. Finally, that the comparative statics in  $a$  are ambiguous, follows from the examples provided in Appendix B.3. **Q.E.D.**

**Proof of Lemma 4.** The assumption in the Lemma is sufficient to ensure that, given a contract as characterized in Theorem 1, the agent's problem  $\max_{\tilde{a}} \{V_A(\tilde{a})\}$  as characterized in (IC) is strictly concave. Hence, the first-order condition in (2) is both necessary and sufficient for incentive compatibility. **Q.E.D.**

**Proof of Lemma 5.** Incentive compatibility and limited liability imply that (PC) must be slack if  $v = 0$ , for any  $\tau \geq 0$  (the agent could always secure himself a payoff of  $V_A \geq 0$  by choosing  $a = 0$ ). Hence, as shown in the main text, the optimal payout time is given by (11) and  $B(\tau) > B(0)$  for  $\tau > T^*(a)$  follows directly from (IC) together with  $I(\tau|a) > I(T^*(a)|a)$ , where we have used Assumption 1. **Q.E.D.**

**Proof of Lemma 6.** See main text. **Q.E.D.**

**Proof of Proposition 3.** For  $\tau = T_{RE}(a|0)$  the implicit function theorem implies

$$\text{sgn} \left( \frac{da_{RE}(\tau)}{d\tau} \right) = \text{sgn} \left( - \frac{\partial^2 W(a|\tau)}{\partial a \partial \tau} \Big|_{\tau=T_{RE}(a|0)} \right) = \text{sgn} \left( \frac{dT_{RE}(a|0)}{da} \right), \quad (17)$$

and the result for marginal regulation follows from the assumption that  $T_{RE}(a|0) = T^*(a)$  is increasing in  $a$  at  $a = a^*$ .<sup>43</sup>

It remains to show that  $\lim_{\tau \rightarrow \infty} a_{RE}(\tau) = 0$ . To do so, write

$$\frac{dW(a|\tau)}{da} = \left[ \frac{1}{I(\tau|a)} \frac{c'(a)}{c'(a)} - \frac{\frac{d}{da} I(\tau|a)}{I^2(\tau|a)} \right] e^{\Delta r \tau} c'(a), \quad (18)$$

and note that, from incentive compatibility, the term in square brackets in (18) is strictly positive for any  $a > 0$ . To see this recall that, given a contract, the agent solves  $\max_{\tilde{a}} \{V_A(\tilde{a})\}$  (cf. (IC)) such that incentive compatibility requires that, at  $\tilde{a} = a$ , both the agent's first-order condition  $B = c'(a)/I(\tau|a)$  as well as the second-order condition

---

<sup>43</sup>In general, the comparative statics in (17) are ambiguous, as illustrated in Corollary B.1 in Appendix B.3 (cf. also Proposition 2).

$B \left( \frac{\partial}{\partial a} I(\tau|a) + I^2(\tau|a) \right) - c''(a) \leq 0$  are satisfied. Substituting out  $B$  we thus must have that

$$\frac{1}{I(\tau|a)} \frac{c''(a)}{c'(a)} - \frac{\frac{\partial}{\partial a} I(\tau|a)}{I^2(\tau|a)} \geq 1. \quad (19)$$

Hence, as  $\frac{d}{da} I(\tau|a) = \frac{\partial}{\partial a} I(\tau|a)$  by the envelope theorem,<sup>44</sup> marginal costs as expressed in (18) go to infinity as  $\tau \rightarrow \infty$  for any  $a > 0$ , and the result follows from (10) and the assumptions on  $\pi(a)$ . **Q.E.D.**

**Proof of Lemma 7.** The following auxiliary Lemma, will be used repeatedly in the remainder of the Appendix:

**Lemma A.2** *The agent's utility under a  $\mathcal{C}_{MI}$ -contract with payout time  $T$ ,  $V_A(a, T) := \frac{c'(a)}{I(T|a)} - c(a)$ , is increasing in  $a$  and strictly decreasing in  $T$ .*

**Proof of Lemma A.2.** The comparative statics of  $V_A$  in  $T$  follow directly from strict monotonicity of  $I(T|a)$  (cf. Lemma 1 and Assumption 1). For the marginal effect of  $a$  consider

$$\frac{\partial V_A(a, T)}{\partial a} = \frac{c''(a) I(T|a) - c'(a) \frac{d}{da} I(T|a)}{I^2(T|a)} - c'(a) = \frac{c'(a)}{I(T|a)} \left( \frac{c''(a)}{c'(a)} - \frac{\frac{d}{da} I(T|a)}{I(T|a)} \right), \quad (20)$$

and the result follows from the arguments in the proof of Proposition 3.  $\blacksquare$

It then follows immediately from  $V_A(a, \tau) = \frac{c'(a)}{I(\tau|a)} - c(a) < v$  that  $a$  cannot be implemented with a  $\mathcal{C}_{MI}$ -contract, and, using Lemma 5, that (PC) must bind. Further, note that  $e^{\Delta r \tau} (v + c(a))$  constitutes a lower bound on wage costs given  $\tau$ , which can clearly only be achieved when all payments occur at  $t = \tau$ . The following contract achieves this bound with two payments at date  $\tau$ : A bonus  $b$  for the realization of  $h_{MI}^\tau$  at date  $\tau$  and an additional unconditional payment. (PC) requires that the present value of these two payments is  $B = (v + c(a))$ . To target (IC) the bonus for history  $h_{MI}^\tau$  is calibrated such that the present value is given by  $\frac{c'(a)}{I(\tau|a)}$ . The unconditional payment is set such that the agent values it at  $v + c(a) - \frac{c'(a)}{I(\tau|a)}$ . From  $I(\tau|a) > \frac{c'(a)}{v+c(a)}$  both payments are non-negative, they clearly satisfy (PC) and (IC) and result in expected wage costs of  $e^{\Delta r \tau} (v + c(a))$ . **Q.E.D.**

<sup>44</sup>Formally, we have, using Assumption 1,

$$\frac{d}{da} I(T|a) = \frac{\partial}{\partial a} \frac{\frac{\partial}{\partial a} L_T(a|h_{MI}^T(a))}{L_T(a|h_{MI}^T(a))} + \frac{\partial}{\partial h} \frac{\frac{\partial}{\partial a} L_T(a|h_{MI}^T(a))}{L_T(a|h_{MI}^T(a))} \frac{\partial h_{MI}^T(a)}{\partial a}$$

where the second term is zero as, from (3),  $h_{MI}^T$  maximizes the likelihood ratio for given  $T$ .

**Proof of Proposition 4.** We will first show that for any  $v > 0$  there exists a finite  $\tilde{\tau}$ , such that, for all  $\tau > \tilde{\tau}$  the principal uses a non- $\mathcal{C}_{MI}$ -contract and implements a strictly decreasing action  $a^*(\tau)$  with  $\lim_{\tau \rightarrow \infty} a^*(\tau) = 0$ . To see this, take  $\tau$  large enough such that the regulatory constraint binds and consider the equilibrium action choice in the relaxed problem ignoring (IC), which is the solution to the following strictly concave problem:

$$\tilde{a}(\tau) = \arg \max_a \pi(a) - e^{\Delta r \tau} (v + c(a)). \quad (21)$$

It is obvious from (21) that  $\tilde{a}(\tau)$  is strictly decreasing in  $\tau$  and approaches zero for  $\tau \rightarrow \infty$ . Now note that  $\tilde{a}(\tau)$  can be implemented in the full problem at same wage costs if and only if  $I(\tau|\tilde{a}(\tau)) \geq \frac{c'(\tilde{a}(\tau))}{v+c(\tilde{a}(\tau))}$ , which is equivalent to  $v \geq \tilde{v}(\tau) := V_A(\tilde{a}(\tau), \tau)$ . The sufficient part follows directly from Lemma 7. To show the necessary part, observe that a contract with  $W(a) = e^{\Delta r \tau} (v + c(a))$  cannot feature any delay beyond  $\tau$  due to differential discounting. However, when  $I(\tau|a) < \frac{c'(a)}{v+c(a)}$ , even the  $\mathcal{C}_{MI}$ -contract with payout time  $\tau$ , which is the contract that provides strongest incentives with date- $\tau$  payments only, requires  $B > v + c(a)$  in order to satisfy (IC), implying  $W(a) > e^{\Delta r \tau} (v + c(a))$ . Hence,  $a^*(\tau) = \tilde{a}(\tau)$  whenever  $v \geq \tilde{v}(\tau)$  and existence of a finite threshold  $\tilde{\tau}$  for any  $v > 0$  then follows from  $\lim_{\tau \rightarrow \infty} \tilde{v}(\tau) = 0$ , where we have used  $\lim_{\tau \rightarrow \infty} \tilde{a}(\tau) = 0$  together with Lemma A.2.

It remains to show that a sufficiently small increase in  $\tau$  above the unconstrained optimal payout time  $T^*(a^*(0))$  induces an increase in the action, i.e.,  $a^*(\tau) > a^*(0)$ . When (PC) is slack without regulation this follows from the arguments in the proof of Lemma 5. So assume that  $v > 0$  and that (PC) binds in the absence of regulation. Then, using Assumption 1, the equilibrium action for  $\tau = 0$  satisfies

$$a^* = \arg \max_a \pi(a) - e^{\Delta_r T_1(a)} (v + c(a)), \quad (22)$$

where  $T_1(a)$  is characterized in Lemma 2. By assumption  $a^*$  is interior and, thus, solves the first-order condition  $\pi'(a^*) - e^{\Delta_r T_1(a^*)} c'(a^*) = \Delta_r T_1'(a^*) e^{\Delta_r T_1(a^*)} (v + c(a^*))$ , where the right-hand side is strictly positive as  $T_1(a)$  is strictly increasing in  $a$  from Lemma A.2. Hence, comparing with the first-order condition characterizing  $\tilde{a}(\tau)$  for  $\tau = T_1(a^*)$ , which from (21) is given by  $\pi'(\tilde{a}) - e^{\Delta_r T_1(a^*)} c'(\tilde{a}) = 0$ , we obtain  $a^* < \tilde{a}(T_1(a^*))$  where we have used strict concavity of the problem in (21). Now, using  $V_A(a^*, T_1(a^*)) = v$  and Lemma A.2,  $a^* < \tilde{a}(T_1(a^*))$  is equivalent to  $T_1(a^*) < \tilde{\tau}$ . Thus, for  $T_1(a^*) \leq \tau \leq \tilde{\tau}$ , both (PC) and (IC) are relevant for compensation costs and the principal uses a  $\mathcal{C}_{MI}$ -contract with payment at  $\tau$ . Then, the equilibrium action is entirely determined from the binding constraints, i.e.,  $a^*(\tau) = \underline{a}(\tau)$  which solves  $V_A(\underline{a}(\tau), \tau) = v$ , and the result follows from

$\frac{\partial a^*(\tau)}{\partial \tau} = \underline{a}'(\tau) > 0$  where we have again used Lemma A.2. **Q.E.D.**

## Appendix B Supplementary material

### B.1 Alternative action spaces

#### B.1.1 Binary action set

Theorem 1 extends one-to-one to a binary action set  $a \in \{a_L, a_H\}$ , such that the technical conditions ensuring validity of the first-order approach (cf. Lemma 4) can be dropped. Let  $a_H$  denote the high-cost action which comes at cost  $c_H$ , and  $a_L$  the low-cost action at cost  $c_L = c_H - \Delta c$ , then, incentive compatibility of  $a_H$  requires that

$$\mathbb{E} \left[ \int_0^{\bar{T}} e^{-rAt} db_t \middle| a_H \right] - \mathbb{E} \left[ \int_0^{\bar{T}} e^{-rAt} db_t \middle| a_L \right] \geq \Delta c. \quad (23)$$

Now, we can define the relevant likelihood ratio for action  $a_H$  as

$$I(t|a_H) := \max_{h^t \in H^t} \frac{L_t(h^t|a_H) - L_t(h^t|a_L)}{L_t(h^t|a_H)}, \quad (24)$$

so that we can apply the tools developed in the main text. In particular, as long as the incentive constraint in (23) is relevant for compensation costs, the optimal contract is as characterized in Theorem 1, adjusting for the new measure of informativeness given in (23) and replacing total costs and marginal costs by  $c_H$  and  $\Delta c$  respectively.

#### B.1.2 Multitask environment

One can even apply our results on the optimal timing of incentive pay to a multi-task environment inspired by Bénabou and Tirole (2016). Here, the agent's first task  $a$  has persistent effects as modelled in the main text, whereas the other task  $q$  has immediately observable effects. In particular, the agent first needs to exert unobservable effort  $q \in [0, 1]$  at associated effort cost  $k(q)$  to create a business opportunity with probability  $q$ . If no opportunity arrives, the game ends. If a business opportunity has arrived, which is immediately observable, the agent then chooses action  $a$  at cost  $c(a, q)$ . We make standard convexity assumptions on  $k(\cdot)$  and  $c(\cdot)$ , and, for simplicity, assume that the agent has a sufficiently low outside option. Since the agent will optimally never receive any compensation if no business opportunity has been created, we may write the agent's

payoff as

$$V_A = q(B - c(a, q)) - k(q).$$

Then, the principal's compensation design problem to implement a given  $a$  and  $q$  is

$$W(a, q) = qB \min_{B, w_t} \int_0^{\bar{T}} e^{\Delta r t} dw_t \quad \text{s.t.}$$

$$B = k'(q) + c(a, q) \quad (\text{ICq})$$

$$B \int_0^{\bar{T}} I(t|a) dw_t = \frac{\partial c(a, q)}{\partial a} \quad (\text{ICa})$$

$$dw_t \geq 0 \quad \forall t \quad (\text{LL})$$

Since  $B$  is fixed for given  $a$  and  $q$ , it is immediate that the solution to this compensation design problem is analogous to Lemma 2 where (ICq) has similar effects as a binding participation constraint.

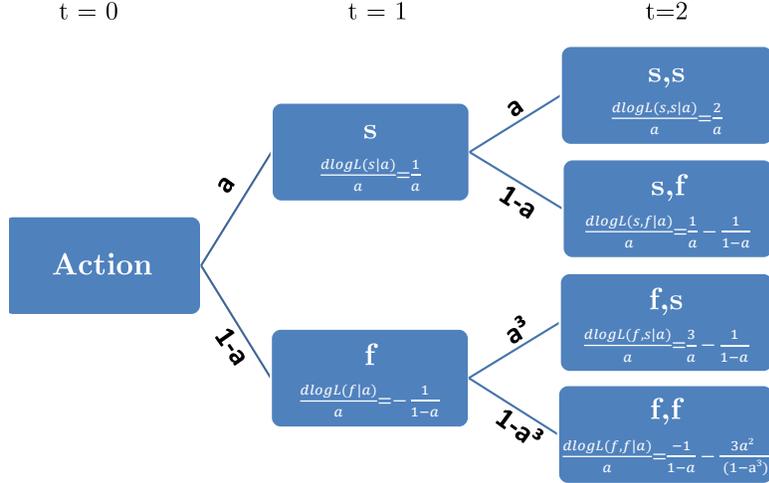
## B.2 Non-i.i.d. example with discrete information arrival

In this Appendix we extend Example 1 in the main text to allow for non-i.i.d. signals. To illustrate the additional insights it is sufficient to again focus on binary signals  $x_t \in \{s, f\}$  and to restrict attention to the two period case, i.e.,  $\bar{T} = 2$ . In particular, consider the information environment depicted in Figure 5 with  $a \in [0, 1]$ . For this concrete specification, a success is the most informative signal in  $t = 1$ , i.e.,  $h_{MI}^1(a) = (s)$  for all  $a$ , while  $h_{MI}^2(a)$  changes with  $a$ : When  $a > \frac{1}{2}$ , the maximum informativeness history at  $t = 2$  is a continuation history of  $h_{MI}^1(a)$ , in particular,  $h_{MI}^2(a) = (s, s)$ . When  $a < \frac{1}{2}$ , however,  $h_{MI}^2 = (f, s)$ , i.e., in this case early failure followed by a success is the best indicator of the agent taking the intended action (cf. e.g., Manso (2011) or Zhu (2016)).

## B.3 Further results for Example 2

In this Appendix we provide some additional material for Example 2 by specifying concrete parametric survival time distributions.

**Example B.1** *Mixed distribution:*  $S(t|a) = aS_L(t) + (1 - a)S_H(t)$  with  $a \in [0, 1]$  and where  $S_L(t)$  dominates  $S_H(t)$  in the hazard rate order, i.e.,  $\lambda_L(t) < \lambda_H(t)$ .



**Figure 5. Example information process.** This graph plots an example information environment with discrete information arrival as in Example 1 but non-i.i.d. signals.

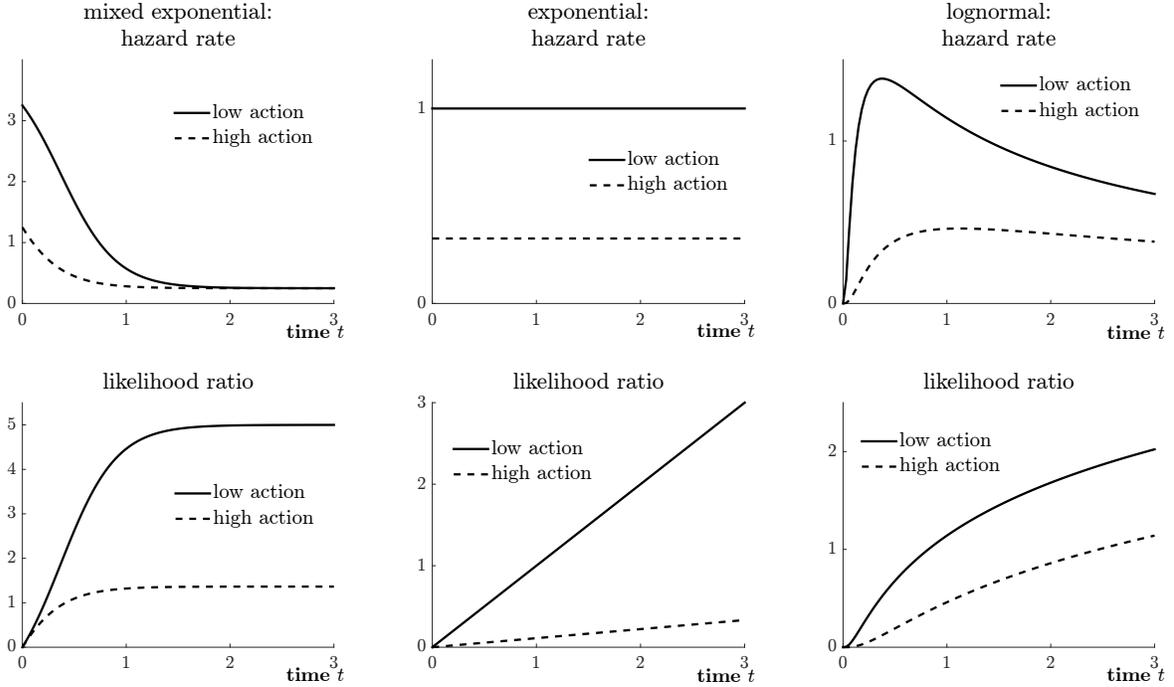
**Example B.2** Weibull distribution:  $S(t|a) = e^{-\left(\frac{t}{a}\right)^\kappa}$ , with  $a > 0$ ,  $\kappa > 0$ .

**Example B.3** Log-normal distribution:  $S(t|a) = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left[ \frac{\log t - a}{\sqrt{2}\sigma} \right]$ , with  $a \geq 0$ ,  $\sigma > 0$ .

Note that we assume that the first-order approach is valid, for which concavity of the survival function  $S$  in  $a$  for all  $(a, t)$  is a sufficient condition. Since this is trivially satisfied for the distribution in Example B.1, which is linear in effort, (see Holmstrom (1984) and Rogerson (1985)), the first-order approach will generally be valid in this parametric example. For the other example distributions, its applicability must be ensured via appropriate parametrization (cf. also Lemma 4).

**Hazard rates, likelihood ratios and effort.** Figure 6 plots the hazard rate  $\lambda(t|a) = -\frac{d \log S(t|a)}{dt}$  and the maximal likelihood ratio  $I(t|a) = \frac{d \log S(t|a)}{da}$  for specifications of Example distributions B.1 to B.3. We plot the functions for two effort levels. It can be inferred from the graph that the slope of  $I$  approaches zero the smaller the difference in hazard rates under high and low effort.

**Optimal payout times.** Since  $I(0|a) = 0$ , optimal contracts are as characterized in Theorem 1. As  $I(t|\cdot)$  is differentiable in this application, the payout time with slack (PC),  $T_{RE}(a)$ , is characterized by the first-order condition in (8). For the Weibull family considered in Example B.2, which nests the exponential distribution for  $\kappa = 1$ , this



**Figure 6. Example information processes.** This graph plots the hazard rate  $\lambda(t|a) = -\frac{d \log S(t|a)}{dt}$  (top panels) and the maximal likelihood ratio  $I(t|a) = \frac{d \log S(t|a)}{da}$  (lower panels) for different arrival time distributions falling in the class of Examples B.1 to B.3 (from left to right). The left panels show a mixed exponential distribution with respective parameters  $\lambda_L = 1/4$  and  $\lambda_H = 4$ , respectively. The middle panels show the case of an exponential distribution (Weibull with  $\kappa = 1$ ) with parameter  $\lambda = 1/a$  and the right panels a lognormal distribution with parameters  $\mu = a$  and  $\sigma = 1$ . For each distribution the plots show the respective functions for two actions corresponding to a mean arrival time of 1 (low action) and 3 (high action), respectively.

trade-off yields a particularly simple closed-form solution  $T_{RE}(a) = \frac{\kappa}{\Delta_r}$ . For the mixed exponential distribution, which is a concrete case of the more general mixed distribution in Example B.1,<sup>45</sup> we obtain  $T_{RE}(a) = \frac{1}{\Delta_\lambda} \log \left( 1 + \frac{\Delta_\lambda - \Delta_r + \sqrt{(\Delta_\lambda - \Delta_r)^2 + 4\Delta_\lambda \Delta_r a}}{2\Delta_r a} \right)$ , where  $\Delta_\lambda := \lambda_H - \lambda_L$ . It is then straightforward to show that, from (9), the payout time  $T_{RE}$  is decreasing in the action in Example B.1 and independent of effort in Example B.2. While there is no closed form solution for  $T_{RE}$  in Example B.3, we can show that, in this case, the optimal payout time is increasing in effort. We summarize these results in the following Corollary:

**Corollary B.1** *Consider the case where (PC) is slack. Then a marginal increase in  $a$  decreases  $T_{RE}(a)$  in Example B.1, has no effect in Example B.2, and increases  $T_{RE}(a)$  in Example B.3.*

<sup>45</sup>Concretely, in this case we have  $S_L(t) = e^{-\lambda_L t}$  and  $S_H(t) = e^{-\lambda_H t}$ , with  $\lambda_H > \lambda_L$ .

When (PC) binds, the maximum number of payments is generally two (see Theorem 1). Since informativeness is continuous in our application (see Figure 6) with  $I(0|a) = 0$ , there always exists a unique payout time  $T_1(a)$  solving  $I(T_1|a) = \frac{c'(a)}{v+c(a)}$  if (PC) binds.<sup>46</sup> Whether this single payout date is optimal depends on how the growth rate of informativeness changes with time. As can be inferred from Figure 6, the exponential distribution has the special feature that informativeness grows linearly. As a result,  $C(I|\cdot)$  is strictly convex and the unique payment date is  $T_1(a) = a^2 \frac{c'(a)}{v+c(a)}$ . In contrast, the other example survival distributions may imply the optimality of an up-front payment at date  $T_S = 0$  and a long run payment at date  $T_L > T_1$  if the agent's outside option is sufficiently high. Again, as a concrete specification of Example B.1, the mixed exponential distribution admits (partial) closed-form solutions. In particular, there exist thresholds  $v_1 < v_2$  such that for  $v > v_2$  two payment dates  $T_S = 0$  and  $T_L(a)$  solving  $\frac{1-e^{\Delta_r T}}{1-e^{\Delta_\lambda T}} \frac{1}{1+a(e^{\Delta_\lambda T}-1)} = \frac{\Delta_r}{\Delta_\lambda}$  are optimal if  $a < \frac{1}{2} \left(1 - \frac{\Delta_r}{\Delta_\lambda}\right)$ , while if the latter condition is violated or the outside option is still sufficiently high,  $v_1 \leq v \leq v_2$ , there is a single optimal payout date  $T_1(a) = \frac{1}{\Delta_\lambda} \log \left(1 + \frac{c'(a)}{v+c(a)-c'(a)a}\right)$ .<sup>47</sup> While we do not obtain closed form solutions for the lognormal distribution in Example B.3, we have verified numerically that, also in this case, it is optimal to specify two payment dates,  $T_S = 0$  and  $T_L > T_1$  for  $\Delta_r$  sufficiently low and  $v$  sufficiently high.

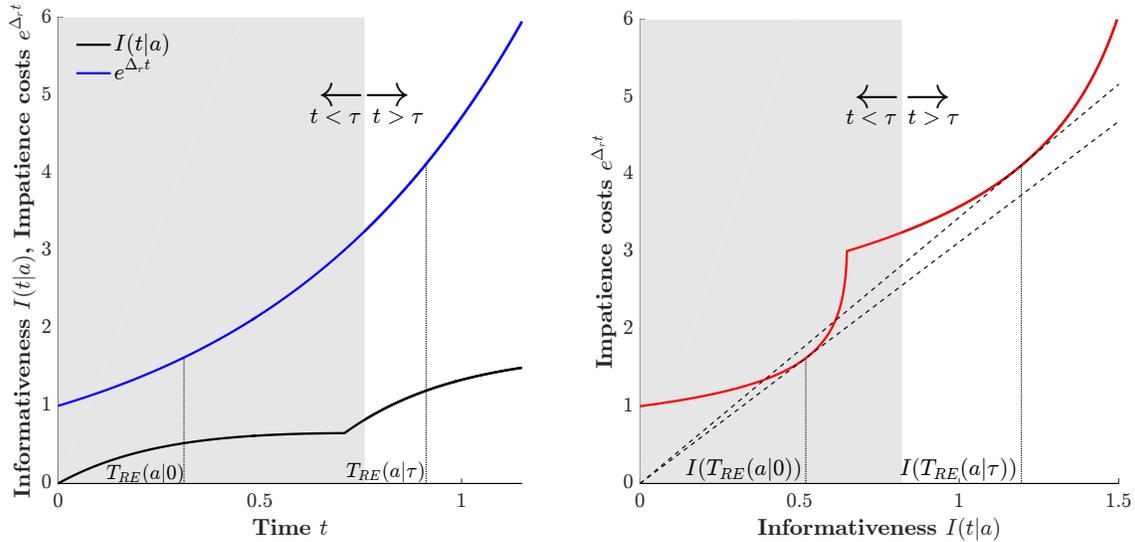
**Deferral regulation without strict convexity of  $C(I|\cdot)$  - An example.** The following parametric specification for Example 2 violates part 2 of Assumption 1 such that a marginal increase in  $\tau$  might lead to a discrete jump in the duration of the optimal contract.

**Example B.4** *Piecewise mixed exponential: The survival time distribution satisfies  $S(t|a) = 1 - wG(t|a)$  for  $t \leq s$  and  $S(t|a) = S(s|a) - wG(t - s|a)$  for  $t > s$ , where  $G(\cdot)$  is mixed exponential, i.e.,  $G(t|a) = aG_L(t) + (1 - a)G_H(t)$  with  $G_L$  and  $G_H$  denoting the cdfs of two exponential distributions with parameters  $\lambda_L$  and  $\lambda_H > \lambda_L$  respectively, and the weight  $w := 1/(1 + G(s|a))$  scaling total probability to 1.*

The left panel of Figure 7 plots the information process corresponding to the piecewise mixed exponential arrival time distribution in Example B.4, which features two phases of high growth. As a consequence, impatience costs as a function of informativeness plotted in the right panel exhibit a non-convex region, implying a linear segment in the

<sup>46</sup> Note, if  $\lim_{t \rightarrow \infty} I(t|a) < \frac{c'(a)}{v+c(a)}$ , then (PC) must be slack for action  $a$ .

<sup>47</sup> If the outside option is below  $v_1$  (PC) is slack. The two thresholds on  $v$  satisfy  $v_1 = c'(a) \left(a + \frac{1}{e^{\Delta_\lambda T_{RE}(a)} - 1}\right) - c(a)$  and  $v_2 = c'(a) \left(a + \frac{1}{e^{\Delta_\lambda T_{LE}(a)} - 1}\right) - c(a)$ , for  $a < \frac{1}{2} \left(1 - \frac{\Delta_r}{\Delta_\lambda}\right)$ .



**Figure 7. Payout times under deferral regulation.** This graph plots informativeness  $I(t|a)$  and impatience costs  $e^{\Delta_r t}$  as a function of time (left panel) as well as impatience costs as a function of informativeness (right panel) for the piecewise mixed exponential arrival time distribution from Example B.4. The parameter values are set to  $\lambda_L = 1/4$ ,  $\lambda_H = 4$ ,  $s = 0.71$ ,  $a = 0.95$  and  $\Delta_r = 1.55$ .

corresponding cost of informativeness. In this example, the optimal unconstrained payout time (pinned down by the tangent line through the origin with the minimum slope,  $\frac{e^{\Delta_r t}}{I(t|a)}$ ) violates the regulatory constraint, i.e.,  $T_{RE}(a|0) < \tau$ . However, the endogenously chosen payout time given the constraint exceeds the minimum deferral period,  $T_{RE}(a|\tau) > \tau$ , so that one may (wrongly) infer that the constraint is irrelevant.

## References

- ALLEN, F., AND D. GALE (2000): “Financial Contagion,” *Journal of Political Economy*, 108(1), 1–33.
- BAUMOL, W. J. (1972): “On Taxation and the Control of Externalities,” *The American Economic Review*, 62(3), 307–322.
- BEBCHUK, L. A., AND J. M. FRIED (2010): “Paying for Long-Term Performance,” *University of Pennsylvania Law Review*, 158, 1915–1960.
- BÉNABOU, R., AND J. TIROLE (2016): “Bonus Culture: Competitive Pay, Screening, and Multitasking,” *Journal of Political Economy*, 124(2), 305–370.

- CARNEY, M. (2014): “The Future of Financial Reform,” 2014 Monetary Authority of Singapore Lecture given by Mark Carney, Chair of the Financial Stability Board.
- CASELLA, G., AND R. L. BERGER (2002): *Statistical inference*, vol. 2. Duxbury Pacific Grove, CA.
- CHAIGNEAU, P., A. EDMANS, AND D. GOTTLIEB (2016): “The Value of Informativeness for Contracting,” Working Paper 20542, National Bureau of Economic Research.
- DEMARZO, P., AND D. DUFFIE (1999): “A Liquidity-based Model of Security Design,” *Econometrica*, 67(1), 65–99.
- DEMARZO, P. M., AND Y. SANNIKOV (2006): “Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model,” *Journal of Finance*, 61(6), 2681–2724.
- DIAMOND, D. W. (1991): “Debt Maturity Structure and Liquidity Risk,” *The Quarterly Journal of Economics*, 106(3), 709–737.
- GJESDAL, F. (1982): “Information and Incentives: The Agency Information Problem,” *The Review of Economic Studies*, 49(3), 373–390.
- GOPALAN, R., T. MILBOURN, F. SONG, AND A. V. THAKOR (2014): “Duration of Executive Compensation,” *The Journal of Finance*, 69(6), 2777–2817.
- GROSSMAN, S. J., AND O. D. HART (1983): “An Analysis of the Principal-Agent Problem,” *Econometrica*, 51(1), 7–45.
- GUTTENTAG, J., AND R. HERRING (1984): “Credit Rationing and Financial Disorder,” *The Journal of Finance*, 39(5), 1359–1382.
- HARRIS, M., AND A. RAVIV (1979): “Optimal incentive contracts with imperfect information,” *Journal of Economic Theory*, 20(2), 231–259.
- HARTMAN-GLASER, B., T. PISKORSKI, AND A. TCHISTYI (2012): “Optimal securitization with moral hazard,” *Journal of Financial Economics*, 104(1), 186 – 202.
- HÉBERT, B. (2015): “Moral Hazard and the Optimality of Debt,” Discussion paper, Stanford University.
- HOFFMANN, F., R. INDERST, AND M. M. OPP (2013): “Regulating deferred incentive pay,” Discussion paper, University of Bonn, University of Frankfurt, University of California, Berkeley.

- HOLMSTROM, B. (1979): “Moral Hazard and Observability,” *Bell Journal of Economics*, 10(1), 74–91.
- (1984): “A Note on the First-Order Approach in Principal Agent Models,” Unpublished working paper, Yale University.
- HOPENHAYN, H., AND A. JARQUE (2010): “Unobservable persistent productivity and long term contracts,” *Review of Economic Dynamics*, 13(2), 333 – 349.
- INNES, R. D. (1990): “Limited liability and incentive contracting with ex-ante action choices,” *Journal of Economic Theory*, 52(1), 45–67.
- JARQUE, A. (2010): “Repeated moral hazard with effort persistence,” *Journal of Economic Theory*, 145(6), 2412–2423.
- JEWITT, I. (1988): “Justifying the First-Order Approach to Principal-Agent Problems,” *Econometrica*, 56(5), 1177–1190.
- JEWITT, I., O. KADAN, AND J. M. SWINKELS (2008): “Moral hazard with bounded payments,” *Journal of Economic Theory*, 143(1), 59 – 82.
- KIM, S. K. (1995): “Efficiency of an Information System in an Agency Model,” *Econometrica*, 63(1), 89–102.
- (1997): “Limited Liability and Bonus Contracts,” *Journal of Economics and Management Strategy*, 6(4), 899–913.
- KUHNEN, C. M., AND J. ZWIEBEL (2009): “Executive pay, hidden compensation and managerial entrenchment,” Unpublished working paper, Princeton University.
- MALAMUD, S., H. RUI, AND A. WHINSTON (2013): “Optimal incentives and securitization of defaultable assets,” *Journal of Financial Economics*, 107(1), 111 – 135.
- MANSO, G. (2011): “Motivating Innovation,” *The Journal of Finance*, 66(5), 1823–1860.
- MERTON, R. (1977): “An analytic derivation of the cost of deposit insurance and loan guarantees An application of modern option pricing theory,” *Journal of Banking and Finance*, 1(1), 3–11.
- MILGROM, P. R. (1981): “Good news and bad news: Representation theorems and applications,” *The Bell Journal of Economics*, pp. 380–391.

- OPP, C. C., M. M. OPP, AND M. HARRIS (2013): “Rating agencies in the face of regulation,” *Journal of Financial Economics*, 108(1), 46–61.
- OPP, M. M., AND J. Y. ZHU (2015): “Impatience versus Incentives,” *Econometrica*, 83(4), 1601–1617.
- PLANTIN, G., AND J. TIROLE (2015): “Marking to market versus taking to market,” Discussion paper, Toulouse School of Economics.
- RAY, D. (2002): “The Time Structure of Self-Enforcing Agreements,” *Econometrica*, 70(2), 547–582.
- RENNER, P., AND K. SCHMEDDERS (2015): “A Polynomial Optimization Approach to Principal-Agent Problems,” *Econometrica*, 83(2).
- ROGERSON, W. P. (1985): “The First-Order Approach to Principal-Agent Problems,” *Econometrica*, 53(6), 1357–67.
- SANNIKOV, Y. (2014): “Moral Hazard and long-run incentives,” Unpublished working paper, Princeton University.
- SHAVELL, S. (1979): “Risk Sharing and Incentives in the Principal and Agent Relationship,” *The Bell Journal of Economics*, 10(1), 55–73.
- THANASSOULIS, J. (2012): “The Case for Intervening in Bankers Pay,” *The Journal of Finance*, 67(3), 849–895.
- TIROLE, J. (2005): *Theory of Corporate Finance*. Princeton University Press, Princeton.
- ZHU, J. (2016): “Myopic Agency,” Unpublished working paper, University of Pennsylvania.