Abstract

We characterize optimal contract design in settings where the principal observes informative signals about the agent's initial action over time. Under bilateral risk-neutrality and limited liability of the agent, all relevant features of a signal process can be encoded in a single “informativeness” function. This function is increasing in time and fully captures the notion of "only time will tell." We then show how the principal uses the timing dimension of the compensation contract to extract rents from the agent via the use of more informative performance signals. Optimal contracts trade off the rent-extraction benefit of deferral with the associated costs resulting from the agent's relative impatience. Our framework lends itself to evaluate the effects of recent regulatory proposals in the financial sector mandating the deferral of bonus payments and the use of claw-back clauses.
Only time will tell
A theory of deferred compensation

Florian Hoffmann† Roman Inderst‡ Marcus Opp§

September 2016

Abstract
We characterize optimal contract design in settings where the principal observes informative signals about the agent’s initial action over time. Under bilateral risk-neutrality and limited liability of the agent, all relevant features of a signal process can be encoded in a single “informativeness” function. This function is increasing in time and fully captures the notion of “only time will tell.” We then show how the principal uses the timing dimension of the compensation contract to extract rents from the agent via the use of more informative performance signals. Optimal contracts trade off the rent-extraction benefit of deferral with the associated costs resulting from the agent’s relative impatience. Our framework lends itself to evaluate the effects of recent regulatory proposals in the financial sector mandating the deferral of bonus payments and the use of claw-back clauses.

Keywords: Compensation design, persistence, principal-agent models, financial regulation.

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

PRELIMINARY AND INCOMPLETE

*The paper subsumes insights from an earlier working paper titled “Regulating deferred incentive pay.” We thank Willie Fuchs, Nicolae Găreanu, Denis Gromb, Ben Hermelin, Gustavo Manso, Robert Marquez, John Morgan, Andy Schwartz, Philipp Strack, Johan Walden, and Jeffrey Zwiebel for valuable insights. Hoffmann and Inderst gratefully acknowledge financial support from the ERC (Advanced Grant “Regulating Retail Finance”).

†University of Bonn. E-mail: fhoffmann@uni-bonn.de.
‡Johann Wolfgang Goethe University Frankfurt. E-mail: inderst@finance.uni-frankfurt.de.
§Harvard University. E-mail: marcus opp@fas.harvard.edu.
1 Introduction

The timing of executive compensation has received widespread attention in the public debate in the aftermath of the financial crisis: Short-term oriented compensation packages are blamed for having contributed to excessive risk-taking in the financial sector, in particular, since large bonuses to financial executives have been paid out before the realization of disastrous outcomes.

“Compensation schemes overvalued the present and heavily discounted the future, encouraging imprudent risk-taking and 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.”


Similar regulatory interventions in the timing of pay have been introduced or proposed around the world.\footnote{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 the UK and has lead to mandatory deferral of bonuses for several years.} While the regulators’ underlying implicit conjectures of 1) “Compensation contracts with short-term bonus payouts induce high risk-taking” and 2) “Mandating longer deferral periods causes less risk-taking” have intuitive appeal, our paper will argue that both conjectures are misguided. The conjectures fail to recognize that the timing of bonus payments in compensation contracts is an equilibrium outcome of the principal-agent relationship between bank shareholders and the bank CEO. Put differently, to account for the Lucas-critique in assessing effects of regulatory intervention, we require a model in which risk-taking and relevant terms of the compensation contract, in particular, the timing of pay, are jointly determined.

Our paper develops such a model of the timing of pay to shed light on how the bank responds to regulatory interference in the timing dimension of optimal compensation contracts, such as a mandated minimum-deferral requirement, by adjusting other dimensions of the compensation contract, i.e., the size and contingency of bonus payments. These contract adjustments are the key reason for why pure deferral regulation tends to backfire, i.e., leads to higher risk-taking. On an abstract level, the issue with deferral regulation is that it only targets a symptom of the bank’s risk-taking incentives, the compensation contracts of its key risk-takers, but does not address the root of these risk-taking incentives, say high leverage. However, our analysis also shows that deferral
regulation is not a complete lost cause in deterring excessive risk-taking as long as it is coupled with additional regulatory restrictions on the contingency of pay, which may be implemented by bonus clawback in the case of bank failure.

While we have motivated the importance of understanding the timing of pay with recent regulatory proposals within the financial sector, one may envision many relevant real-life settings in which the timing of pay plays a crucial role. For example, the quality of a CEO’s strategic decision is usually only observed with delay, the true value of an innovative activity of a R&D unit can often be assessed only well into the future, and, whether a manager of a private equity or real estate fund put effort in selecting the most promising investments is only observed once these illiquid investments are sold. All of these settings share the important feature that the agent’s action has persistent effects. Thus, while our main application focuses on the financial sector, where consequences of many actions, such as exposures to tail risks, are not immediately observable and may only be revealed in downturns, the general contract design implications of persistence we identify have wider applicability. From this perspective, other than speaking to topical issues in financial regulation, our analysis also makes a more general contribution to the theory of incentive compensation.

Our general compensation design setup builds on canonical static principal-agent environments, such as Holmstrom (1979) or Harris and Raviv (1979), but we suppose that the principal receives informative signals about the agent’s initial action over time. Conditioning compensation contracts on informative long-run signals is costly since liquidity needs make the agent effectively impatient relative to the principal (see e.g., DeMarzo and Duffie (1999)). For the case of bilateral risk-neutrality and limited liability of the agent (cf. Innes (1990), Park (1995) and Kim (1997)), we provide a complete characterization of the optimal timing and contingency of compensation.

Initially, consider the optimal contingency of compensation to implement a given action. If a contract stipulates a bonus payment for a particular history at a given point in time $t$, then this history must maximize the likelihood ratio over all possible date-$t$ histories.$^2$ The intuition is akin to a static environment. Due to risk-neutrality, there are no risk-sharing considerations, so that the agent is only rewarded for this “best possible” history in order to provide maximal incentives. Different from a static environment, the construction of this best possible history for each $t$ implies that the associated informativeness is now an increasing function of time.$^3$ Optimal payout times are designed to trade-off this informativeness benefit of deferral with the costs resulting

---

$^2$Technically, this requires the action to be implementable with this history.

$^3$Intuitively, the principal could always ignore newly arriving information.
from agent impatience. We show that, regardless of the information environment, there are at max two payout dates. In particular, if the agent’s outside option is sufficiently low so that she receives an agency rent, there is a single payout time which reflects both the principal’s rent extraction motive as well as (social) impatience costs. If the size of the compensation package is determined by the outside option, the timing solely aims to minimize impatience costs, and we show that this may result in two payout dates depending on how informativeness increases over time.

To make predictions above and beyond optimal compensation design, we focus on an application to the financial sector and interpret the agent’s action as costly effort which reduces the hazard rate of bank failure. Hence, lack of effort corresponds to more “risk-taking.” Roughly speaking, this is supposed to capture that generating “true alpha” is costly to the agent, while taking on (down-side) risk is not. Compensation contracts are designed according to the just discussed general optimality principles. In particular, to implement a given effort level at lowest cost, the bank rewards its risk-taker only after the “best possible” history, i.e., in the absence of bank failure. Interestingly, the optimally chosen payout time(s) may then both increase as well as decrease in risk-taking, depending on the survival time distribution.

While the just described optimal contracts minimize compensation costs for any given effort level, the induced equilibrium effort may imply excessive risk-taking from society’s perspective, for instance, as a bank’s board or shareholders (the principal) may not fully internalize the social cost of bank failure. In fact, the bank optimally induces the effort level that maximizes the present value of the associated gross profit stream net of the associated minimum compensation costs. Since society cares about bank failure above and beyond the bank’s own incentives (as captured in the bank’s valuation of the gross profit stream), our setup features an additional external agency problem between the bank and society: Society prefers a lower level of risk-taking (higher effort) than the one implemented by the bank which provides scope for regulatory intervention. Concretely, we study the effects of deferral regulation on the effort level induced by the bank, and hence, the equilibrium frequency of bank failure. Such analysis has been called for by the Financial Stability Board, a quasi-regulatory body for financial institutions.

“The effectiveness of these mechanisms remains largely untested and more analysis is needed to assess whether tools such as malus and clawbacks are sufficiently developed and effectively used to deter risks.”


It is important to note, that such a regulatory intervention does not really address the
source of excessive risk-taking incentives, i.e., the external agency problem, but instead intervenes in the internal agency problem between the bank and the bank CEO. However, there may be good reasons, such as limited liability of bank shareholders, that prevent (or constrain) the trivial solution of having bank shareholders directly internalize the externalities from bank failure. In these cases, thus, may be a role for second-best regulation such as deferral regulation. Still, the results from our analysis are not supportive of pure deferral regulation.

To understand how deferral regulation typically goes wrong, one has to understand the incentives of the bank. The bank induces an effort level that equates the bank’s marginal gross profits to its marginal compensation costs. Since regulation only constrains compensation design, but not affect the gross profits, regulation is effective at lowering risk-taking (increasing effort) if and only if it reduces marginal compensation cost. However, there is a robust effect that causes mandatory deferral to pull towards an increase in marginal compensation cost.

To see this, consider, first, the case where the agent has a low outside option. By optimality of the unconstrained payout time, any required deferral above and beyond this payout time causes the impatience costs of delay to grow faster than the associated gain in informativeness. (Otherwise, the principal would have chosen to wait for more precise information himself.) Importantly, this effect applies for every dollar that the agent receives in present value terms, robustly pulling towards an increase in marginal (and, trivially, the level of) compensation cost. In addition to this unambiguous “timing-inefficiency” effect, the total effect of mandatory deferral on marginal costs also depends on how the (growth rate of the) informativeness changes in effort itself. For small deviations from the principal’s unconstrained payout time, the former “timing-inefficiency” effect is still second-order, and we demonstrate that regulation may either increase or decrease risk-taking depending on the survival time distribution. However, for sufficiently stringent regulation the induced timing inefficiency always dominates leading robustly to higher risk-taking. Intuitively, the principal reacts to the intervention in the timing of pay, by reducing the size of the payment package.

This argument no longer applies for the case of a sufficiently high outside option. When the size of the agent’s compensation package is fixed by her outside option and the principal is forced to defer sufficiently, making the entire pay contingent on the most informative history (absence of bank failure) generates such strong incentives that it would become impossible to implement low effort levels. While this looks like “good news,” the principal now “optimally” adjusts the compensation contract by paying out part of the compensation package unconditionally, i.e., regardless of bank failure. Such adjust-
ments obviously countervail the underlying regulatory intention, but may be effectively constrained by additional “clawback requirements” based on bank failure. The optimal design of these clawback clauses requires that all payments be contingent on the most informative history, the absence of bank failure. Thus, a suitable combination of deferral regulation and clawback provisions may then be effective in mitigating risk-taking and increasing welfare.

In short, jointly the recently implemented regulatory measures of deferral regulation and clawback clauses can be successful in reducing risk-taking if the agent’s outside option is high. Else, the principal is able to adjust other relevant margins of the compensation contract in order to “contract around regulation,” which, typically, results in higher rather than lower risk-taking.

**Literature.** Our basic agency problem is similar to standard static moral-hazard settings with risk-neutral agent and limited liability such as Innes (1990), Park (1995) and Kim (1997). While these classic contributions assume that performance measures are immediately observable, we incorporate the notion of “only time will tell,” i.e., the principal observes signals which are imperfectly correlated with the agent’s (initial) action over time. For general information environments all the information relevant for compensation design under bilateral risk-neutrality can be encoded into an increasing “informativeness” function of time, which allows for a tractable analysis of the optimal timing of pay, determined by the trade-off between a more informative performance measure and the agent’s relative impatience. Our model thus nests the important example of moral hazard by a securitizer of defaultable assets, as studied in Hartman-Glaser, Piskorski, and Tchistyi (2012) and Malamud, Rui, and Whinston (2013).

Hopenhayn and Jarque (2010) also study general information environments in a moral-hazard problem with persistence, but consider the case of a risk-averse agent (cf. also the repeated action models in Jarque (2010) and Samnikov (2014)). Due to the agent’s desire to smooth consumption (across time and states), risk-aversion generates different trade-offs in compensation design, which, e.g., do not allow to collapse the information environment into a simple increasing function of time. In our model, this representation, however, is key to capture the informativeness benefit of deferral in a particularly tractable way.

That the timing of pay determines the information about the agent’s hidden action that the principal can use for incentive compensation further relates our analysis to the

---

4 Innes (1990) considers limited liability constraints for both the agent and the principal, in our model, as in Park (1995) and Kim (1997), only the agent is protected by limited liability.
more general problem of comparing information systems in agency problems. There is a large literature deriving 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)). While in our model longer delay unambiguously generates “more informativeness” as captured by an increase in the likelihood ratio, it also generates endogenous costs due to the wedge in discount rates. The optimal contract optimally resolves the resulting trade-off, thereby determining the optimal information structure in equilibrium.\(^5\)

Recently, there has been increasing interest in theories, like ours, that analyze and motivate regulatory interference in (bankers’) pay, even in the absence of internal governance failures. While this literature is mostly concerned with restrictions on the size of compensation (e.g., Thanassoulis (2012) and Bénabou and Tirole (2016))\(^6\) our focus is on regulatory interventions in the timing dimension of pay\(^7\).

**Organization of paper.** Our paper is organized as follows. In Section 2 we present our basic agency problem of persistent moral hazard and solve the compensation design problem, i.e., characterize explicitly the compensation cost minimizing contract for a given abstract action, including the optimal contingency and timing of pay. In Section 3 we then present our main application to the financial sector, for which we analyze the complete problem of the principal consisting of optimal compensation design under deferral regulation and the equilibrium action choice. Section 4 concludes. Proofs not contained in the main text are in Appendix A. Appendix B contains some supplementary material.

## 2 General compensation design model

### 2.1 Setup

Our general modeling framework is a moral hazard problem with bilateral risk-neutrality and limited liability of the agent in which the principal observes informative signals about

\(^5\)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.

\(^6\)For a related analysis of a moral hazard problem with lower and upper bounds on compensation without a regulatory motivation see Jewitt, Kadan, and Swinkels (2008).

\(^7\)See also Inderst and Pfeil (2013) for timing regulation in a two-period multi-task model when there is an inherent conflict of interest between loan generation and screening.
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 abstract action (the first problem in Grossman and Hart (1983)). Our application to the finance sector in Section 3 then also considers the optimal action choice.

Time is continuous \( t \in [0, \bar{T}] \). At time 0, an agent \( A \) with outside option \( v \) takes an unobservable, one-time 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 action \( a \) affects the distribution of a stochastic process of verifiable signals \( X(t) \) that may arrive continuously (see application in Section 3) or at discrete points in time (as in the “toy”-environment in Figure 1). The abstract signals may refer to output realizations, annual performance reviews by the principal, or multidimensional objects consisting of all informative signals. The respective date-\( t \) history of realized signals \( h^t = \{ x_s \}_{s=0}^t \in H^t \) then captures all relevant information about the agent’s action available to the principal at time \( t \).

![Figure 1. Example information process.](image)

This graph plots an example information environment. The agent takes the action at date 0, but no information is revealed until date \( t = 1 \). Then, the realized signal is either success or failure, i.e., \( x_1 \in \{ S, F \} \). At \( t = T = 2 \), the signal is either milestone, success or failure, i.e., \( x_2 \in \{ M, S, F \} \). The probabilities for each path are a function of the initial action \( a \).

A contract \( \mathcal{C} \) stipulates transfers from the principal to the agent as a function of the information available at the time of payouts. Formally, we capture these transfers through the cumulative compensation process, \( b(t) \) adapted to the filtration \( \mathcal{F}_t \) 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,

---

8 Formally, the index set \( S \) of the stochastic process \( X(t) \) can be any compact subset of \([0, \bar{T}]\).

9 See Holmstrom (1979) for a definition of informative signals.
it is costly to do so since we assume that the agent is relatively impatient. That is, the
discount rates of the agent, $r_A$, and the principal, $r_P$, satisfy

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

Relative impatience of the agent is a standard assumption in dynamic principal-agent
models (see e.g., DeMarzo and Duffie (1999), DeMarzo and Sannikov (2006), or Opp and
Zhu (2015)). It is typically motivated with liquidity needs on the side of the agent.

### 2.2 Compensation design problem

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

**Problem 1 (Compensation design)**

$$W(a) := \min_{b(\cdot)} \mathbb{E}\left[\int_0^T e^{-r_P t} db(t) \bigg| a\right] \quad \text{s.t.}$$

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

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

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

The first constraint is the agent’s participation constraint [PC]. 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 $b(t)$. Limited
liability of the agent [LL] requires $db(t) \geq 0$. Since signals do not necessarily correspond
to output realizations defining the principal’s gross payoff, in contrast to Innes (1990),
we neither impose limited liability of the principal nor a monotonicity constraint.

10 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.

10 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.
that the first-order approach is valid. (We present sufficient technical conditions for its validity in Section 2.4, where we also show that our results readily extend to a binary action set). Hence, for each $a$, we can replace (IC) by the following first-order condition

$$\frac{\partial}{\partial a} \mathbb{E} \left[ \int_0^T e^{-rA_t} db(t) \right] = c'(a). \tag{2}$$

For ease of exposition, our subsequent analysis makes use of three auxiliary variables. First, let $B$ denote the agent’s time-0 valuation of the compensation package (using her discount rate), or henceforth, the size of the compensation package

$$B := \mathbb{E} \left[ \int_0^T e^{-rA_t} db(t) \right]. \tag{3}$$

Second, let $w(s)$ denote the fraction of the compensation package that the agent derives from stipulated payouts up to time $s$, i.e.,

$$w(s) := \mathbb{E} \left[ \int_0^s e^{-rA_t} db(t) \right] / B, \tag{4}$$

so that $w(T) = \int_0^T dw(t) = 1$. One may then interpret $\int_0^T tdw(t)$ as the duration of the compensation contract. Third, if we scale the marginal incentives provided by the contract, $\frac{\partial}{\partial a} B$ (the left hand side of (2)) with the size of the compensation package, we obtain a measure for contract informativeness, $L_\mathcal{C}$, capturing the incentives provided per unit of expected pay

$$L_\mathcal{C} := \frac{\mathbb{E} \left[ \int_0^T e^{-rA_t} db(t) \right]}{B}. \tag{5}$$

The units of $L_\mathcal{C}$ can formally be interpreted as a weighted average log likelihood ratio of performance signals used in the compensation contract $\mathcal{C}$. Intuitively, if a contract only stipulates an unconditional payment, then $L_\mathcal{C} = 0$, since the compensation package does not vary with $a$.

We can then express the objective function of the principal as $W(a) = B \int_0^T e^{\Delta_t} dw(t)$. Thus, the wage costs are simply the product of the size of the agent’s compensation package, $B$, and the weighted average (gross) impatience costs $\int_0^T e^{\Delta_t} dw(t)$. Now, from (2), incentive compatibility implies that $BL_\mathcal{C} = c'(a)$, i.e., the size of the compensation package is inversely related to contract informativeness. As a consequence, the optimal

---

11 In static models, the problems associated with this approach were first pointed out in Mirrlees (1974) and Grossman and Hart (1983). Sufficient conditions for the applicability of the first-order approach are given in Rogerson (1985) and Jewitt (1988).
contract will usually use only the most informative performance signals for a given payout date \( t \). In fact, as we will show below, the only instance where this does not hold is the (uninteresting) case without a relevant agency problem. Our analysis, thus, proceeds in three steps. We first characterize the properties of these maximal-incentives contracts and then analyze the optimal timing of payouts within this class. Finally, we give a precise condition for when the optimal contract is within the class of maximal-incentives contracts and characterize optimal contracts when this condition does not hold.

2.3 Analysis

2.3.1 Maximal-incentives contracts

To draw on basic insights from static moral hazard models, first suppose that the payout time \( t \) was exogenously fixed. Since impatience costs, \( e^{\Delta r t} \), are also fixed for a given payout time \( t \), optimal contract design is then akin to a static model. In static principal-agent models with bilateral risk-neutrality and limited liability of the agent maximal-incentives contracts are optimal. Due to the absence of risk-sharing considerations, the agent is only rewarded for the outcome with the highest likelihood ratio regarding the recommended action (and obtains zero rewards for all other outcomes due to limited liability).

Our preliminary goal is to define a natural analog to these maximal-incentives contracts in our dynamic setting. As a first step, this requires us to define the most informative outcome, i.e., the most informative history of signals, \( h^t_{MI}(a) \), for a given \( t \). To abstract from technical considerations in defining this history, let us initially suppose that, for all \((t, a)\), each history \( h^t \) occurs with strictly positive ex-ante probability mass, \( \Pr(h^t|a) \), and, that \( \Pr(h^t|a) \) is differentiable in \( a \) (see Section 2.4 for generalizations). Then, \( h^t_{MI} \) maximizes the log likelihood ratio over all possible date-\( t \) histories \( H^t \)

\[
h^t_{MI}(a) := \arg \max_{h^t \in H^t} \frac{\partial}{\partial a} \frac{\Pr(h^t|a)}{\Pr(h^t|a)}
\]

For example, in the toy environment shown in Figure 1, a conventional specification would imply that the most informative history at time 1 is success in \( t = 1 \), i.e., \( h^1_{MI}(a) = (S) \) and a sequence of a success and a milestone by date 2, i.e., \( h^2_{MI}(a) = (S, M) \)

---

12 In this case all payments are made at \( t = 0 \) such that we are in fact back to a static model.
13 See e.g., Innes (1990) or, for a textbook treatment, Tirole (2005).
14 It is possible that multiple histories maximize this likelihood ratio. It is then without loss of generality to pick any of these histories.
15 An example specification is: \( \Pr(x_1 = S|a) = 1 - \Pr(x_1 = F|a) = a \) and \( \Pr(x_2 = M|a) = \frac{a}{2} \).
With slight abuse of notation, let $db(t|h^t)$ denote the contractually stipulated bonus payout at time $t$ that the agent receives for a realization of history $h^t \in H^t$. Then,

**Definition 1** A maximal-incentives contract $C_{MI}$ never stipulates agent rewards for any history other than $h^t_{MI}$ histories, i.e., $db(t|h^t) = 0$ for all $t$ whenever $h^t \neq h^t_{MI}$.

The following Lemma highlights the key role that $C_{MI}$-contracts play in our analysis.

**Lemma 1** If it is possible to implement action $a$ with a $C_{MI}$-contract, then the optimal contract is a $C_{MI}$-contract.

The proof of this central Lemma argues by contradiction and shows how contracts outside the class of $C_{MI}$-contracts can be perturbed to yield lower wage costs. Suppose the optimal contract stipulated payouts at time $t$ for some history $h^t \neq h^t_{MI}$ and \( (PC) \) was slack under this contract, then the principal could lower the size of the compensation package $B$ in an incentive-compatible way by shifting rewards towards more informative histories (holding the payout time fixed), leading to a contradiction. Second, suppose \( (PC) \) binds and the optimal contract stipulated payouts for some history $h^t \neq h^t_{MI}$ where $t > 0$. The principal can now maintain incentive compatibility by lowering the fraction of the compensation package deferred until date $t$ and pay out a sufficient amount at an earlier date to satisfy \( (PC) \). Similarly, suppose a payout was stipulated for some history $h^0 \neq h^0_{MI}$ and the contract also provided additional incentives with stipulated payouts for $t > 0$, then the principal could increase the fraction of pay derived from date-0 payouts, $w(0)$, and correspondingly reduce the fraction of pay by future bonus payments while maintaining incentive compatibility. In both cases, the perturbations keep the agent’s compensation fixed and strictly lower the weighted average impatience costs. This reduces the principal’s wage cost $W$, leading again to a contradiction.

The main idea behind the proof already sheds light on the circumstances when $C_{MI}$-contracts do not apply, in which case \( (PC) \) binds and all compensation is paid out at date 0, i.e., $w(0) = 1$ (see exact condition in Theorem 1). Since the timing of pay is thus only relevant when $C_{MI}$-contracts apply, it is only essential to characterize optimal payout times within this class of contracts.

$\Pr(x_2 = S|a) = \frac{1}{2}$, and $\Pr(x_2 = F|a) = \frac{1}{2} - \frac{a}{2}$. Of course, by varying the specification of the information process, it may be possible that early failure is an indicator of taking the intended long-run action (see e.g., Manso (2011) or Zhu (2015)). That is, the most informative history at date 2, as given by (6), then would be an initial failure followed by a milestone $(F, M)$. 

11
2.3.2 Optimal payout times

The principal’s choice of payout times $t$ can be interpreted as choosing the quality of the information environment. This optimization over the timing dimension is absent in static environments. Specifically, in our example environment in Figure 1 we have to determine whether the principal should optimally stipulate rewards for date 0, 1, 2, or, any combination of the three. Intuitively, optimal payout times are pinned down by the trade-off between impatience costs, measured by $e^{\Delta r_t}$, and gains in informativeness. Since $\mathcal{C}_{MI}$-contracts only stipulate rewards for maximally informative histories, date-$t$ informativeness is appropriately measured by the likelihood ratio for action $a$ evaluated at history $h^t_{MI}(a)$.

**Lemma 2** Maximal informativeness $L(t|a)$ is an increasing function of time

$$L(t|a) := \frac{\partial}{\partial a} \Pr (h^t_{MI}|a) \Pr (h^t_{MI}|a).$$

(7)

Contract informativeness of $\mathcal{C}_{MI}$-contracts satisfies

$$L_{\mathcal{C}} = \int_0^T L(t|a) dw(t).$$

(8)

Since $L(t|a)$ traces out the maximal likelihood ratio over all possible histories, $L(t|a)$ must be an increasing function of time (see left panel of Figure 2 for an example). 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. Since $\mathcal{C}_{MI}$-contracts only make use of these histories, the associated contract informativeness is just a weighted average of the (maximal) informativeness $L(t|a)$ at the respective payout dates (see (8)).

To determine the optimal timing of $\mathcal{C}_{MI}$-contracts it is now convenient to transform Problem 1 into a deterministic problem by viewing $B$ and $w(\cdot)$ as control variables (rather than $b(\cdot)$).

**Problem 2**

$$W(a) = \min_{B,w(\cdot)} B \int_0^T e^{\Delta r_t} dw(t)$$

16 Given $B$ and $w(t)$, one may readily solve for $db(t)$ from $\mathbb{E}[db(t)|a]e^{-r_A t} = Bd w(t)$.  

12
Figure 2. Impatience costs versus informativeness: The left panel plots both impatience costs, $e^{\Delta r t}$, and maximal informativeness, $L(t|a)$, 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 $L_C$ for all possible contracts. The solid circles in the right panel correspond to the boundary points that define the (minimum) cost of informativeness $C(L_C|a)$.

\[
B \geq v + c(a) \quad \text{(PC)}
\]
\[
B \int_0^T L(t|a)dw(t) = c'(a) \quad \text{(IC)}
\]
\[
dw(t) \geq 0 \quad \forall t \quad \text{and} \quad w(T) = 1 \quad \text{(LL)}
\]

The transformed problem clearly shows that as long as two signal generating processes feature the same maximal informativeness, $L(t|a)$, the timing of optimal contracts will be identical (even if likelihood ratios differ along less informative histories). From now on, one may thus think of the function $L(t|a)$ (rather than $X(t)$) as the primitive of the information environment that fully captures the formalization of “time will tell.” The timing of pay is then determined by the trade-off between weighted average impatience costs, $\int_0^T e^{\Delta r t}dw(t)$, and contract informativeness, $\int_0^T L(t|a)dw(t)$. The shaded region in the right panel of Figure 2 depicts all feasible combinations of average impatience costs and contract informativeness $L_C$, including contracts outside the class of $\mathcal{CMI}$-contracts. Of particular relevance for the subsequent analysis is its lower boundary $C(L_C|a)$.

Of particular relevance for the subsequent analysis is its lower boundary $C(L_C|a)$.

\[
\text{17 All feasible combinations can be constructed as the convex hull of $\mathcal{CMI}$-contracts with single payout dates $\{ (L(t|a), e^{\Delta r t}) \}_{t \in [0,T]}$, and, in addition, the 2 points (0, 1) and $(0, e^{\Delta r T})$. The latter two points are defined by uninformative contracts that stipulate an unconditional payment at date 0 and $T$, respectively.}
\]
Definition 2 The cost of informativeness, $C(L_{\mathcal{E}}|a)$, is the minimal impatience cost for a given level of contract informativeness $L_{\mathcal{E}}$.

By construction, $C$ is an increasing and convex function mapping $L_{\mathcal{E}} \in [0, L(T|a)]$ into $[1, e^{\Delta r \cdot T}]$. Since it is always possible to exclusively pay out at date $t$, $C(L(t|a)|a)$ is bounded above by $e^{\Delta r \cdot t}$, i.e., $C(L(t|a)|a) \leq e^{\Delta r \cdot t} \forall t$. Given the cost of informativeness, we are now ready to analyze the optimal timing of pay.

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 (due to the inverse relationship between $B$ and $L_{\mathcal{E}}$ in (IC)). 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

$$T_{RE}(a) = \arg \min_t e^{\Delta r \cdot t} |_{t = T_{RE}(a)} \geq 18$$

which implies contract informativeness of $L_{\mathcal{E}} = L(T_{RE}(a)|a)$. Graphically, it can be identified by the point on $(L_{\mathcal{E}}, C(L_{\mathcal{E}}|a))$ that minimizes the slope of a ray through the origin. From the right panel of Figure 2 we can immediately deduce that $T_{RE}(a) = t^*$ in this example. Moreover, if we assume differentiability of $L$ in $t$, we can characterize $T_{RE}(a)$ in terms of an intuitive first-order condition

$$\frac{d \log L}{dt} \bigg|_{t = T_{RE}(a)} = \Delta r. \quad (10)$$

That is, the principal defers until the (log) growth rate of informativeness, $\frac{d \log L}{dt}$, 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 choice of payout times minimizes the contract’s impatience costs, $\int_0^T e^{\Delta r \cdot t} dw(t)$, subject to ensuring incentive compatibility which requires from (IC) that contract informativeness satisfies $L_{\mathcal{E}} = \int_0^T L(t|a) dw(t) = \frac{c'(a)}{v + c(a)}$. Since minimal impatience costs $C \left( \frac{c'(a)}{v + c(a)} \right) |a)$

\[18\] 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 $V_A = \frac{c'(a)}{L(T_{RE}(a)|a)} - c(a)$, while the principal is indifferent.

\[19\] Section 3 covers a relevant example with differentiable $L$. 14
can be achieved by using a convex combination of at most two payout dates, this puts a tight upper bound on the number of payout dates specified by optimal contracts with binding (PC).

**Lemma 3** 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 $L(T_1|a) = \frac{c'(a)}{v + c(a)}$ and, given a solution, this payout date achieves minimal impatience costs, i.e., $e^{\Delta r T_1(a)} = C \left( \frac{c'(a)}{v + c(a)} \right) a$.

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

![Figure 3. PC binds one vs. two dates](image)

**Figure 3. PC binds one vs. two dates:** The left panel plots the case when $C$ is a strictly convex function.

Lemma 3 implies that there are generically two payment dates for discrete information processes, since then there is typically no single date $T_1(a)$ that solves $L(T_1|a) = \frac{c'(a)}{v + c(a)}$. For illustrative purposes, it is thus instructive to explain the economics behind the choice of one versus two payment dates in an environment where $L$ is continuous. In the left panel of Figure 3, $C$ is strictly convex so any convex combination of two $\mathcal{C}_{MI}$-contracts is inefficient. In particular, this implies suboptimality of a natural candidate contract

---

20 In general, these two payout dates are unique (as in Figure 3). In some knife-edge cases, there may be multiple solutions for $T_S$ and $T_L$. In this knife-edge case, it is without loss of generality to pick 2 payout dates.
indicated by the red circle: This (inefficient) contract supplements the optimal contract in the case of slack (PC), which pays out at date $T_{RE}(a)$, with a sufficient up-front payment to satisfy (PC).

In contrast, in the right panel of Figure 3, two payout dates are optimal as they exploit non-convexities in the impatience costs of 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 2). 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).

Theorem 1 Suppose $L(0|a) \leq \frac{c'(a)}{v+c(a)}$, then action $a$ is optimally implemented with a $CMI$-contract.

1) If $v \leq \bar{v} = \frac{c'(a)}{L(T_{RE}|a)} - c(a)$, (PC) is slack, the unique payout date is $T_{RE}(a)$ and the size of the compensation package is $B = \frac{c'(a)}{L(T_{RE}|a)}$.

2) Otherwise, (PC) binds, so that $B = v + c(a)$ and $L_a = \frac{c'(a)}{v+c(a)}$. The choice of payout dates is as characterized in Lemma 3.

Given this characterization the associated wage cost to the principal, the solution to Problem follows immediately.

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

(11)

Theorem summarizes the surprisingly simple characterization of $CMI$-contracts in general information environments. The trade off between informativeness and impatience costs is optimally resolved with at max two payment dates. Only when (PC) is slack, the choice of the payout time is influenced by a rent extraction motive of the principal.

It remains to explicitly characterize optimal contracts in the (less interesting) case when $CMI$-contracts do not apply.

Lemma 4 If $L(0|a) > \frac{c'(a)}{v+c(a)}$, $CMI$-contracts do not apply. Then, (PC) must bind and all payments are made at time 0, so that $W(a) = B = v + c(a)$.

Intuitively, if the principal receives sufficiently precise signals at time 0, our model (with $v > 0$) generates the same results as a static model with sufficiently informative
signals. In fact, in this case, the agent’s compensation is pinned down by her outside option, $B = v + c(a)$, and the incentive constraint, (IC), becomes irrelevant for compensation costs (see Park (1995) and Kim (1997)). To understand the role of the threshold value $L(0|a) = \frac{c'(a)}{v + c(a)}$, 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, $h^0_{MI}$. Then, the marginal benefit of increasing the action to the agent is $L(0|a) (v + c(a))$. If this exceeds the marginal cost, $c'(a)$, $C_{MI}$-contracts provide excessive incentives. In this case, the optimal contract is not unique. The principal can implement the action at minimal costs, for example, by making a fraction $\gamma_{MI}$ of the pay package contingent on $h^0_{MI}$ and pay the remainder $(1 - \gamma_{MI}) B$ as an unconditional up-front transfer.\(^\text{21}\) Here, $\gamma_{MI}$ is set such that

$$\gamma_{MI} = \frac{c'(a)}{L(0|a) (v + c(a))}. \quad (12)$$

We have now generally characterized the optimal timing and contingency of compensation contracts to implement any given action $a$. For ease of exposition, we have made a few strong assumptions that allowed us to abstract from technical issues. The purpose of the following section is to highlight the generality of the ideas presented in this section, and show that most technical issues within our framework have natural counterparts to those in standard static environments.

### 2.4 Robustness

#### 2.4.1 Signal structure

In this section, we show how our framework extends to the natural case of an uncountable set of possible histories $H^t$ as well as an infinite horizon $T = \infty$. For each given $(a, t)$, consider, thus, the probability space $(H^t, \mathcal{F}^t, \mu^t(\cdot|a))$, where $\mu^t(\cdot|a)$ denotes the probability measure assigned to elements of the filtration $\mathcal{F}^t$ as a function of $a$. The relevant condition that defines maximal-incentives histories $h^t_{MI}(a)$, and hence $L(t|a) = \frac{\frac{\partial}{\partial a} \mu(h^t_{MI}(a)|a)}{\mu(h^t_{MI}(a)|a)}$,\(^\text{21}\) Take the extreme case of $a = 0$ and $v > 0$, then the principal would want to make an unconditional payment of $v$ to provide zero incentives.
then becomes

\[ h^t_{MI}(a) := \arg \max_{h^t \in H^t} \frac{\partial}{\partial a} \mu(h^t|a) \cdot \mu(h^t|a). \]  

(13)

Then, in order for our characterization of optimal compensation contracts for a given action \( a \) (cf. Theorem 1) to extend to this generalized environment, we require the following technical conditions:

**Assumption 1** Sufficient technical conditions:

1. For each \( t \), \( \mu(h^t|a) \) is differentiable in \( a \) for all possible histories \( h^t \).
2. A maximizer in (13), i.e., \( h^t_{MI}(a) \), exists for all dates \( t \).
3. The first-order approach is valid. A sufficient condition is that \( \mu(h^t_{MI}(a)|\tilde{a}) \) is strictly concave in \( \tilde{a} \).

Each of these conditions rules out technical issues that similarly arise in static models. The differentiability condition (1) is a necessary ingredient to define (13). To see the issue that Condition (2) addresses, 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 \( a \) is increasing in the signal realization of \( x_0 \) (such as under MLRP), then \( h^0_{MI}(a) \) does not even exist in this static environment. Finally, our analysis is based on the assumption that the first-order condition for the agent-optimal choice of \( a \) is also sufficient for incentive compatibility. This is ensured whenever the agent’s problem is strictly concave. A sufficient condition is strict concavity of \( \mu(h^t_{MI}(a)|\tilde{a}) \) in \( \tilde{a} \) which is reminiscent of the convexity of the distribution function condition (CDFC) in static models (cf. e.g. Rogerson (1985)).

Finally, to ensure that optimal contracts as characterized in Theorem 1 are economically meaningful, we impose the following additional restrictions on the environment:

**Assumption 2** Economic relevance of setup:

(a) For any finite \( t \), \( L(t|a) \) is bounded above and \( \lim_{t \to \infty} \frac{L(t|a)}{e^{\Delta rt}} = 0 \).

(b) The probability mass of history \( h^t_{MI}(a) \) is strictly positive for any finite \( t \).

Condition (a) can be interpreted as “slow learning.” It is akin to assuming the absence of perfectly revealing signals in static models and ensures that the incentive problem is

\[ 22 \text{When } H^t \text{ is countable, using our previous notation, we can write } \mu(h^t|a) = \Pr(h^t|a) \text{ and directly obtain (6). Similarly, consider the case where there is a continuum of possible histories that can be mapped one-to-one into an interval, } H^t = (a^t, b^t) \subseteq \mathbb{R}, \text{ and suppose that } \mu^t \text{ is absolutely continuous, then, in a slight abuse of notation, } \mu^t(h^t|a) = f(h^t|a)dt, \text{ where } f(\cdot|a) \text{ is a (univariate) density function. The extension to the multivariate case follows analogously. Further, our notation is general enough to also capture measures that have both absolutely continuous parts as well as mass points.} \]

\[ 23 \text{In fact the static case is nested in our model for } T = 0. \]
relevant for compensation design. The condition also implies that the principal would never choose to defer until \( t = \infty \), i.e., impatience costs are relevant. Condition (b) implies that all stipulated bonus payments to the agent are bounded. If history \( h_{MI}^{t} (a) \) had zero mass, an infinite bonus payment needs to be made to satisfy (IC). While the (expected) size of the compensation package \( B \) would still be bounded, such incentive-schemes would push the implications of risk-neutrality to the extreme. Taken together, if Assumptions 1 and 2 are satisfied for the implemented action \( a \), then the optimal contract is characterized by Theorem 1 and stipulates bounded bonuses at finite payout dates.

### 2.4.2 Action space

Differentiability and the validity of the first-order approach might seem restrictive. Hence, it is worthwhile to point out that Theorem 1 extends one-to-one to a binary action set, so that Conditions (1) and (3) of Assumption 1 can be dropped. Let \( a_H \) denote the high-cost action \( a_H \) at cost \( c_H \) and \( a_L \) denote the low-cost action at cost \( c_L = c_H - \Delta c \), then, incentive compatibility of \( a_H \) requires that

\[
E \left[ \int_0^T e^{-r A t} db (t) \middle| a_H \right] - E \left[ \int_0^T e^{-r A t} db (t) \middle| a_L \right] \geq \Delta c
\]

(14)

Now, we can define the relevant likelihood ratio for action \( a_H \) as

\[
L(t|a_H) := \max_{h^t} \frac{\mu(h^t|a_H) - \mu(h^t|a_L)}{\mu(h^t|a_H)},
\]

(15)

so that we can apply the tools developed in the previous section. In fact, we show in the Appendix that one can even apply our results on the timing of optimal incentives contracts to a multi-task environment inspired by Bénabou and Tirole (2016). Here, the agent’s first task \( a \) has persistent effects, whereas the other task \( q \) has immediately observable effects (see Appendix B.1 for details).

### 3 Application to financial sector

#### 3.1 Economic and regulatory environment

In recent years, the timing dimension of (executive) compensation has come under increased regulatory scrutiny. As discussed in the introduction, various countries now mandate a minimum deferral time of bonuses paid out to executives and key risk takers in the financial sector. The heuristic rationale behind this kind of regulation is that, if
regulation remains longer at risk and can be withheld in case of the realization of large risks such as the bankruptcy of the whole institution, this should provide more incentives to act with diligence and, hence, reduce the risk of bank failure. However, mandating longer deferral than what would be privately optimal, restricts the contracting space and, thus, imposes additional costs on banks. In order to analyze the effects of this kind of regulatory intervention we now explicitly specify how the agent’s effort affects the performance signals available for contract design, and, further, add a payoff function specifying how the agent’s action maps into the principal’s profit. This additional structure allows us to study not only optimal compensation design in the presence of deferral regulation, but also the equilibrium action, i.e., the question of which action the principal will optimally induce taking into account possible regulatory constraints.

The agent should now be interpreted as a CEO or key risk-taker of a financial institution and the principal as (the board of) a bank. We assume that the agent’s action, to be interpreted as costly effort, controls the arrival time distribution function \( F(t|a) \) of a verifiable bad event which we interpret as bank failure.\(^{[24]}\) In this setting, the performance signal available to the principal at date \( t \) is the counting process, \( X(t) \), where \( X(t) = 1 \) indicates that bank failure has occurred before time \( t \), and \( X(t) = 0 \) otherwise.

We suppose that the survival function, \( S(t|a) = 1 - F(t|a) = Pr(X(t) = 0|a) \) is twice continuously differentiable in both arguments. Then, to capture the notion that higher effort reduces the instance of bank failure, we posit:

Assumption 3 The hazard rate \( \lambda(t|a) = -\frac{d\log S(t|a)}{dt} \) is strictly decreasing in \( a \) for all \( t \in (0, \infty) \):

\[
\frac{\partial \lambda(t|a)}{\partial a} < 0 \quad \forall t \in (0, \infty), a.
\] (16)

Assumption \(^{[24]}\) plays a similar role as the monotone likelihood ratio property (MLRP) in static principal-agent models with immediately observable signals, see e.g., \(^{[1981]}\) \(^{[20]}\). It is satisfied for our three leading example survival distributions.

Example 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., \( \lambda_L(t) < \lambda_H(t) \).

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

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

\(^{[24]}\) While we focus on the application to the financial sector, one can, more generally interpret the bad event as a disaster, such as a critical accident or a nuclear meltdown.

\(^{[20]}\) It implies a first-order stochastic dominance order on the (family of) survival distributions indexed by \( a \), i.e., \( \partial S/\partial a \geq 0 \).
We continue to assume that the first-order approach is valid. In this application to the financial sector, concavity of the survival function \( S \) in \( a \) for all \((a,t)\) is a sufficient condition for its validity.\(^{26}\)

Effort generates a (bounded) payoff to the principal denoted by \( \pi(a) \), which should be interpreted as the present value of the (gross) profit streams. Take the example of an exponential survival distribution (Example 2 with \( \kappa = 1 \)) and suppose that the bank generates a flow fee income of \( f \) until bank failure, then \( \pi(a) = \frac{f}{r_{p+a}} \). More generally, we assume that \( \pi(a) \) is strictly increasing and weakly concave in \( a \), which means that the bank benefits from higher effort (ignoring compensation costs). However, due to a negative externality of bank failure on society, the bank does not bear the full social cost of failure. We model this wedge between society’s and the banks’s objectives in reduced form and assume that the present value of this externality, \( x(a) \), is decreasing in \( a \) and that \( \pi(a) - x(a) \) is concave. One may think of many reasons why bankers’ objectives are not perfectly aligned with society’s goals, e.g., bank shareholders exploiting the regulatory guarantees implied by deposit insurance (see, e.g., Merton (1977)), externalities of bank failure on other banks (see, e.g., Allen and Gale (2000)), or a shorter time horizon of the bank’s board than the one of society (see, e.g., Guttentag and Herring (1984)).\(^{27}\)

The goal of our analysis is to study the effects of (minimum) deferral regulation that restricts compensation to be paid out on or after date \( \tau \), i.e., \( b(t) = 0 \quad \forall t < \tau \).\(^{28}\) We focus on this single regulatory tool because of its practical relevance and the current debate about its effectiveness. While such an adhoc restriction on the regulatory toolset does not allow us to solve for the optimal regulation for a given friction, it does allow us to analyze the narrower question of whether regulation has the intended effect of raising effort incentives. The benefit of this narrower scope, essentially a comparative statics analysis, is that we can remain agnostic about the exact frictions that give rise to the externality \( x(a) \). The reason is that for a fixed action \( a \), deferral regulation does not directly increase the extent to which the principal internalizes the externality, i.e., \( \pi \) and \( x \) are not a function of \( \tau \). Instead, regulation operates only indirectly via the wage cost function \( W \). (This contrasts with minimum capital ratio requirements that affect the wedge between shareholders’ preferences and the one of society, but do not impose direct restrictions on the compensation design space, i.e., \( W(a) \) is unaffected.)

\(^{26}\) Since concavity is trivially satisfied for the distribution in Example 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.

\(^{27}\) Alternatively, the wedge could result from a corporate governance problem, in which the board’s and shareholders’ preferences are not perfectly aligned (see, e.g., Kuhnen and Zwiebel (2009)).

\(^{28}\) The unregulated case is thus nested by setting \( \tau = 0 \).
Formal problem. Following Grossman and Hart (1983), we decompose the analysis of the Principal’s Problem into compensation design under deferral regulation (Problem 3) and optimal action choice (Problem 4).

Problem 3 (Compensation design)

\[
W(a|\tau) = \min_{b(\cdot)} \mathbb{E} \left[ \int_0^\infty e^{-rP} db(t) \right] \quad \text{s.t. } (IC), (PC), (LL)
\]

\[
b(t) = 0 \quad \forall t < \tau \quad \text{(R)}
\]

Problem 4 (Equilibrium action)

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

We define the resulting welfare as:

\[
\omega(\tau) = \pi(a^*(\tau)) - x(a^*(\tau)) - c(a^*(\tau)) - B \int_{\tau}^\infty (e^{\Delta r t} - 1) dw(t)
\]

where the final term measures the deadweight cost of deferred compensation resulting from relative impatience of the agent. First-best welfare (absent an agency problem) does not involve any inefficient delay and the associated first-best action \(a^{FB}\) solves the first-order condition \(\pi'(a^{FB}) - x'(a^{FB}) = c'(a^{FB})\). It is then easy to show that a necessary condition for deferral regulation to be welfare increasing relative to the “laissez-faire” outcome, \(\omega(\tau) > \omega(0)\), is that it is successful in increasing the equilibrium action.\(^{30}\)

3.2 Compensation design

3.2.1 Mapping into general model

To apply the compensation design results of the general model in Section 2 to our concrete application, it is necessary to characterize the maximal-incentives histories \(h_{MI}^t(a)\). This allows us to map the signal process \(X\) into the informativeness function \(L\).

\(^{29}\) We thus make the simplifying assumption that the principal and agent have the same welfare weight. (Note that a possibly different welfare weight on the externality can be incorporated in \(x(a)\).)

\(^{30}\) The intuition is as follows: The principal’s payoff cannot increase with tighter restrictions on the contracting space, and the agent’s payoff does either not depend on \(\tau\) and \(a\) (when (PC) binds) or decreases in \(\tau\) and increases in \(a\) (when (PC) is slack). Combined with \(x'(a) < 0\), the result follows.

---

22
Lemma 5 For any action $a > 0$, the likelihood ratio maximizing history at each date $t$ is a history without previous failure, as summarized by $X(t) = 0$. Hence,

$$\Pr(h_t^t|a) = S(t|a),$$

$$L(t|a) = \frac{\partial S(t|a)/\partial a}{S(t|a)}. \quad (19)$$

Informativeness satisfies $L(0|a) = 0$ and is strictly increasing and differentiable in $t$ with

$$\frac{\partial L(t|a)}{\partial t} = -\frac{\partial \lambda(t|a)}{\partial a} > 0 \quad \forall t \in (0, \infty), a. \quad (20)$$

Since survival by time $t$ implies absence of failure for any $s \leq t$ we obtain that $X(t) = 0$ is a sufficient statistic for a history without bank failure. Intuitively, by our assumption on the hazard rate (Assumption 3), this history is most informative about the action $a$, yielding (19) and (20). Thus, for any finite $t$, the most informative history has positive probability mass (see discussion in Section 2.4). Moreover, 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, $L(0|a) = 0$. Finally, informativeness grows at a faster rate, as measured by $\partial L(t|a)/\partial t$, whenever the hazard rate is more sensitive to effort. To see the intuition for this, think of approximating the local incentive constraint with binary effort. If the hazard rate under high and low effort is identical at time $t$, the principal learns nothing from the absence (or occurrence) of a disaster. If instead, the hazard rate under low effort is much higher than under high effort, the principal learns “a lot.” Formally, this is captured by the term $-\partial \lambda(t|a)/\partial a$, which is strictly positive by Assumption 3.

It is worthwhile to highlight that the structure imposed by our application is, apart from motivating $\pi(a)$ and $x(a)$, only used to express $L$ in terms of primitives. Thus, our results on the effects of deferral regulation on equilibrium effort will continue to hold in more general settings. The special features of $L$ in this setting, differentiability of $L$ and $L(0|a) = 0$ (see Lemma 5), facilitate exposition, but are not directly relevant for the effect of deferral regulation. The latter feature, i.e., the absence of relevant performance signals at date 0, $L(0|a) = 0 \leq \frac{c'(a)}{v+c(a)}$, implies that the principal optimally implements all effort levels with $C_{MI}$-contracts in the absence of regulation: Optimal compensation design in the absence of regulation ($\tau = 0$), in particular the timing of payouts, is completely characterized by our general Theorem 1. For sake of completeness, Appendix B.2 lists the concrete characterization of optimal payout times for the informativeness functions $L$ implied by our example distributions (see plots in Appendix-Figure 5). In particular,
we show that it is often possible to obtain closed-form expressions for payout times and
determine exact conditions for whether a distribution implies one or two payment dates
when (PC) binds.

In light of the regulatory discussion it is interesting to highlight one feature of optimal
compensation design for the case where payout times reflect a rent-extraction motive, i.e.,
(PC) is slack. Then, higher effort may imply earlier or later payout times, depending
on the information environment (here, the arrival time distribution). More formally,
using the first-order conditions in (10), the sign of the comparative statics of payout
times $T_{RE}(a)$ in the implemented effort level depends on whether the (log) growth rate
of informativeness, $\frac{d\log L}{dt}$, increases or decreases in $a$, i.e.,

$$\text{sgn} \left( \frac{dT_{RE}(a)}{da} \right) = \text{sgn} \left( \frac{\partial}{\partial a} \frac{d\log L}{dt} \bigg|_{t=T_{RE}(a)} \right).$$

(22)

The standard examples reveal that all comparative statics are generically possible so
that it is misguided to conclude that short-term payout times are necessarily reflective of
poor incentives, i.e., low effort $a$. The payout time is decreasing in effort in Example 1,
independent of effort in Example 2, and increasing in effort in Example 3.

We now turn to the effect of regulatory deferral on the principal’s (constrained) op-
timal implementation of a given action $a$.

### 3.2.2 Contract design under deferral regulation

This section analyzes which dimensions of the compensation contract are adjusted by the
principal when regulation constrains the timing dimension. The type of adjustments to
implement a given action $a$ depend on whether (PC) binds or not. The key difference is
that when (PC) is slack, the principal can adjust the size of the compensation package
$B$ in response to increases in $\tau$. Otherwise, $B$ is fixed at $v + c(a)$.

**No relevant participation constraint.** It is instructive to first consider the relaxed
problem without a participation constraint, which, e.g., always applies when $v \leq 0$.
Then, regulation constrains the principal’s optimal choice of payout times as soon as
$\tau > T_{RE}(a)$. We denote the optimal payout time given the deferral requirement $\tau$ as

$$T_{RE}(a|\tau) = \arg \min_{t \geq \tau} \frac{e^{\Delta r t}}{L(t|a)}.$$

(23)
so that $T_{RE}(a|0) = T_{RE}(a)$. One may conjecture that $T_{RE}(a|\tau) = \max\{T_{RE}(a), \tau\}$, but, this conjecture is not generally valid. In particular, when $C(L_{\ell}|\cdot)$ is not strictly convex, the bank may choose an optimal payout time of $T_{RE}(a|\tau) > \tau$ when facing a minimum deferral requirement of $\tau$ years, even though it would have chosen $T_{RE}(a) < \tau$ in the absence of regulation (see Example 4 and Figure 6 in Appendix B.2). Going forward, for ease of exposition, we will make an assumption that rules out this case.

**Assumption 4** The cost of informativeness in the absence of regulation, $C(L_{\ell}|a)$, is strictly convex.

**Lemma 6** Consider the relaxed problem without (PC) and suppose the principal is constrained by regulation, i.e., $\tau > T_{RE}(a)$. Then, the optimal contract is a $C_{MI}$-contract with payout time $\tau$ and an adjusted size of the compensation package of

$$B(\tau) = \frac{c'(a)}{L(\tau|a)} \leq B(0) = \frac{c'(a)}{L(T_{RE}(a)|a)}.$$ 

Thus, as soon as the regulatory constraint interferes with the principal’s unconstrained choice, he uses the associated increase in maximal informativeness at the later payout date, $L(\tau|a) > L(T_{RE}(a)|a)$, to reduce the size of the agent’s compensation package $B$. The reduction in $B$ is calibrated to maintain incentive compatibility of the action $a$.

**Relevant participation constraint.** To compactly express when (PC) binds, it is useful to denote the agent’s utility under a $C_{MI}$-contract with a single payout time of $T$ by $V_{A}(a,T)$ which has the following properties:

**Lemma 7** The agent’s utility $V_{A}(a,T) := \frac{c'(a)}{L(T|a)} - c(a)$ is increasing in $a$ and strictly decreasing in $T$.

The comparative statics of $V_{A}$ in $a$ are implied by agent optimality while the comparative statics in $T$ follow from the fact that informativeness is strictly increasing over time. Using this notation, (PC) is thus slack whenever $V_{A}(a,T_{RE}(a|\tau)) > v$ and binding otherwise.

Together with Lemma 6, Lemma 7 implies that a sufficiently stringent deferral period $\tau$ may decrease the size of the compensation package $B$ by so much that (PC) binds under deferral regulation, even if (PC) was slack in the absence of regulation. The associated

---

31 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.
threshold level for \( \tau \) is then characterized by \( V_A(a, \tau) = v \), which can be equivalently stated in terms of informativeness at date \( \tau \), i.e., \( L(\tau|a) = \frac{c'(a)}{v+c(a)} \).

Next consider the case, when \([PC]\) binds even in the absence of regulation, i.e., \( V_A(a, T_{RE}(a)) < v \). Then, the principal chooses a single payout date \( T_1 \) satisfying \( L(T_1|a) = \frac{c'(a)}{v+c(a)} \) in the absence of regulation (by strict convexity of \( C \) and Lemma 3). Regulation thus constrains the principal’s choice of compensation design as soon as \( \tau > T_1 \), or equivalently, \( L(\tau|a) > \frac{c'(a)}{v+c(a)} \).

Taken together, irrespective of whether \([PC]\) is slack or binding in the absence of regulation, we obtain

**Lemma 8** Suppose regulation is so stringent such that \( L(\tau|a) > \frac{c'(a)}{v+c(a)} \), then \([PC]\) binds and regulation prohibits the principal from implementing action \( a \) with a \( C_{MI} \)-contract. All payouts are made at time \( \tau \) implying that \( W(a) = e^{\Delta r \tau} (v + c(a)) \).

Lemma 8 is essentially a Corollary to Lemma 4. If at the earliest payout time \( \tau \), sufficiently informative performance signals are available, \( L(\tau|a) > \frac{c'(a)}{v+c(a)} \), the incentive compatibility constraint becomes irrelevant for compensation costs. The principal then optimally pays out as early as allowed by regulation to economize on impatience costs so that \( W(a) = e^{\Delta r \tau} (v + c(a)) \). Analogously to (12), he can implement the action by making only a fraction \( \gamma_{MI} = \frac{c'(a)}{L(\tau|a)(v+c(a))} \) contingent on the absence of failure and pay out the remainder unconditionally at time \( \tau \).

A common feature of the cases of slack and binding \([PC]\) is that deferral regulation induces the bank to adjust other dimensions of the compensation contract in order to implement a given effort level. As long as \([PC]\) is slack, the bank adjusts only the size of the compensation package while still using \( C_{MI} \)-contracts. If \([PC]\) binds and regulation is sufficiently stringent the bank responds by adjusting the contingency of payouts and, hence, deviates from \( C_{MI} \)-contracts. Understanding these adjustments is crucial to analyze the effect of deferral regulation on the equilibrium effort level, which we will analyze next.

### 3.3 Equilibrium effort

#### 3.3.1 Characterization

As is well known, characterizing the optimal effort level is typically a complex problem even in standard environments. In our setting, this analysis is further complicated by the regulatory constraint. For ease of exposition, we thus follow the structure of our compensation design analysis and first study the profit-maximizing action choice in the
absence of a relevant (PC), i.e.,

\[ a_{RE} (\tau) = \arg \max_a \pi (a) - W_{RE} (a|\tau), \quad (24) \]

where \( W_{RE} (a|\tau) = e^{r T_{RE}(a|\tau)} c'(a) \) denotes the compensation cost function of this relaxed problem and \( T_{RE} (a|\tau) = \max \{ T_{RE} (a), \tau \} \). To facilitate exposition, we require that \( c'(a) \) is sufficiently convex as to ensure strict convexity of \( W_{RE} \). Then, by concavity of \( \pi \), effort \( a_{RE} (\tau) \) is the unique solution to:

\[ \pi'(a_{RE} (\tau)) = \left. \frac{\partial W_{RE} (a|\tau)}{\partial a} \right|_{a=a_{RE}(\tau)}. \quad (25) \]

Let \( v_{RE} (\tau) := V_A (a_{RE} (\tau), T_{RE} (a_{RE} (\tau)|\tau)) \geq 0 \) denote the agent’s utility associated with \( a_{RE} (\tau) \), then, we have the following result:

Lemma 9  Equilibrium effort satisfies

\[ a^*(\tau) = \begin{cases} 
a_{RE} (\tau) & v \leq v_{RE} (\tau) 
\arg \max_a \{ \pi (a) - (v + c (a)) e^{r \max\{\tau, T_{1}(a)\}} \} & v > v_{RE} (\tau) \end{cases}, \quad (26) \]

As is intuitive, when the agent’s outside option is sufficiently small, the participation constraint will be slack at the optimal effort choice, while it will bind for \( v > v_{RE} (\tau) \). In the former case, equilibrium effort, as determined from (24), reflects the principal’s rent-extraction motive, while, in the latter case, it minimizes the costs of deferral, due to the agent’s relative impatience, subject to the constraints imposed by the agency problem and regulation. Based on the characterization of optimal effort in Lemma 9 we will next study the impact of deferral regulation on the equilibrium effort choice.

3.3.2 Comparative statics of minimum deferral regulation

As the previous characterization suggests, the dependence of \( a^*(\tau) \) on the minimum deferral period \( \tau \) will crucially depend on which set of constraints binds in equilibrium. Hence, we will again first study the comparative statics of \( a^*(\tau) \) in \( \tau \) for the case without a relevant participation constraint and, then, second, the case when (PC) binds. Finally, we illustrate, using an example, how the equilibrium effort transitions between these two regimes as \( \tau \) increases from zero to infinity and provide a discussion of the robust effects of deferral regulation.

---

32 As in static models, validity of the first-order approach does not imply that the principal’s problem is quasiconcave in the action (see Grossman and Hart (1983) or Jewitt, Kadan, and Swinkels (2008)).
No relevant participation constraint. When there is no relevant participation constraint, e.g., as \( v \leq 0 \), the equilibrium action is always pinned down by the first-order condition \((25)\). Now, since deferral regulation only affects the principal’s cost of inducing effort, \( W_{RE}(a|\tau) \), but not its benefit, \( \pi(a) \), the impact of deferral on equilibrium effort is solely determined by the effect on the marginal compensation cost \( \frac{\partial W_{RE}(a|\tau)}{\partial a} \). An increase in marginal compensation costs decreases equilibrium effort, while a reduction causes effort to increase.

To understand the effects of minimum-deferral regulation, it is useful to first analyze the (trivial) effect on the level of compensation costs, \( W_{RE}(a|\tau) \). As soon as regulation constrains the principal, \( \tau > T_{RE} \), compensation costs must increase relative to the absence of regulation:

\[
W_{RE}(a|\tau) - W_{RE}(a|0) = c'(a) \left[ \frac{e^{\Delta_{\tau}}}{L(\tau|a)} - \arg\min_{t} \frac{e^{\Delta_{t}}}{L(t|a)} \right] > 0.
\]

Generically, any cost minimization problem cannot produce lower costs if one imposes additional constraints. The severity of the cost increase induced by minimum deferral regulation depends on the difference between \( \frac{e^{\Delta_{\tau}}}{L(\tau|a)} \) and \( \arg\min_{t} \frac{e^{\Delta_{t}}}{L(t|a)} \). Intuitively, while infinitesimal deviations around the principal’s privately optimal timing choice produce second-order effects due to the envelope theorem, the induced timing inefficiencies will eventually become first-order and cause wage costs to explode.

While an increase in the level of compensation costs is, thus, a generic effect of imposing an additional constraint in the compensation design problem, the effect on marginal compensation costs depends critically on the nature of the chosen regulatory intervention. To illustrate this, it is useful to rewrite the constrained wage cost function as \( W_{RE}(a|\tau) = c'(a) \cdot w(\tau|a) \) where \( w(t|a) := \frac{e^{\Delta_{t}}}{L(t|a)} \). The marginal compensation costs can then be decomposed into two terms

\[
\frac{\partial W_{RE}(a|\tau)}{\partial a} = c''(a) w(\tau|a) + c'(a) \frac{\partial w}{\partial a}.
\]

The first term in \((27)\) captures the (direct) effect that it is marginally more expensive to induce higher effort from the agent (due to convexity of her cost function \( c(a) \)) and treats \( w \) as a constant of proportionality. In contrast, the second term accounts for the possibility that the informativeness of performance signals, \( L \), (and, hence, \( w \)) may also depend on the effort level \( a \).

Let us initially analyze the effect of deferral regulation on the first term. Since a multiplicative constant has always the same effect on marginal costs as on the level of costs, the effect of binding deferral regulation on the first term is unambiguous. By optimality
of the principal’s unconstrained timing choice, the forced increase in contract informativeness induced by deferral regulation must be outweighed by the increase in impatience costs, i.e., $c''(a) w(\tau|a) > c''(a) \arg \min_t w(t|a)$. Deferral regulation thus unambiguously pulls this term towards an increase in marginal costs, and, hence, a reduction in equilibrium effort. The key feature of minimum deferral regulation underlying this unambiguous effect is that its costs - resulting from the agent’s relative impatience - are proportional to the size of the compensation package $B$, i.e., deferral regulation increases the principal’s cost for every dollar that the agent receives in present value terms. (If the principals implemented $a = 0$, deferral regulation would impose no costs). It is this specific interaction of the costs of deferral with the agency problem that sets it apart from other investments in a more precise information system (cf. e.g. Gjesdal (1982)).

To highlight the special nature of deferral regulation (above and beyond imposing a constraint on compensation design) consider the following, related regulatory intervention: Suppose the regulator forced the principal to make an additional investment of $k$ in the information system, such as an upgrade of the IT or accounting system, which increases the informativeness of performance signals by a factor of $1 + \Delta_L$ for all actions and all $t$. Then, the total compensation costs (including investment expenditures) are given by $\frac{W_{RE}(a|0)}{1+\Delta_L} + k$. (It is easy to see that for a sufficiently high $k$, the principal would not choose to make this investment in the absence of regulation.) In contrast to deferral regulation, however, such an upgrade of the IT system does not make it more expensive to pay an additional dollar to the agent. Once implemented, its costs are constant and sunk. It is then easy to see that this variant of a forced investment in a more precise information system is effective in reducing the principal’s marginal cost of inducing a higher action, $\frac{\partial W_{RE}(a|0)}{\partial a} \frac{1}{1+\Delta_L} < \frac{\partial W_{RE}(a|0)}{\partial a}$.

For simplicity, we have assumed that the informativeness benefit of an investment in this alternative information system, $\Delta_L$, does not vary with the action $a$. If we imposed an analogous restriction in the case of deferral regulation, i.e., $\frac{\partial w}{\partial a} = 0$, the marginal costs decomposition (27) would immediately imply that deferral regulation unambiguously leads to an increase in marginal costs. However, in general, investments in a more precise information system might have a differential impact on informativeness depending on the induced effort.\(^\text{33}\) In particular, for the case of deferral regulation, consider the second component of marginal costs in (27), $c'(a) \frac{\partial w}{\partial a}$, which precisely captures whether the principal obtains more precise performance signals under low or high levels of effort. Clearly, the concrete information environment dictates which case obtains. Its sign is therefore generally ambiguous, $\frac{\partial w}{\partial a} \leq 0$, and so is the effect of deferral on it $\frac{\partial}{\partial \tau} \frac{\partial w}{\partial a} \leq 0$. This

\(^{33}\) See Chaigneau, Edmans, and Gottlieb (2016) for related ideas.
ambiguous effect, resulting from a possibly differential increase in informativeness across actions, might either reinforce or weaken the unambiguously positive effect of deferral regulation on marginal costs resulting from higher impatience costs. In particular, for infinitesimal interventions in the optimal timing choice, $\tau = T_{RE}(a|0)$, the effect of deferral regulation on the first term in (27) is second-order by the envelope theorem, so that only the ambiguous effect on the second term remains. In contrast, the following Proposition shows, that for large interventions the effect on the first term will eventually dominate.

**Proposition 1** Consider the case with (PC) slack. Then, the effect of binding deferral regulation on the equilibrium action is ambiguous for sufficiently small interventions. In particular, for $\tau = T_{RE}(a|0)$ we obtain:

$$\text{sgn}\left(\frac{da_{RE}(\tau)}{d\tau}\right) = \text{sgn}\left(\frac{dT_{RE}(a|0)}{da}\right) = \text{sgn}\left(\frac{\partial}{\partial a} \frac{d \log L}{dt} \bigg|_{t=\tau}\right).$$

For $\tau$ sufficiently high, the action satisfies $a_{RE}(\tau) < a_{RE}(0)$ with $\lim_{\tau \to \infty} a_{RE}(\tau) = 0$.

This Proposition highlights that sufficiently stringent deferral regulation will always “backfire” in the sense of reducing equilibrium towards zero. Since our arguments did not rely on the specific structure of our application to the financial sector, this negative effect of sufficiently stringent deferral regulation, as well as the other effects discussed above, apply more generally. The additional value of our application is that it allows us to interpret lower effort in terms of higher equilibrium frequencies of disasters in the financial sector. Moreover, we can use our concrete example distributions to pin down the generally ambiguous effect for small regulatory interventions.

**Corollary 1** Consider the case where (PC) is slack in the absence of regulation and let $\tau = T_{RE}(a|0)$. Then a marginal increase in $\tau$ decreases effort in Example 1, has no effect in Example 2, and increases effort in Example 3.

Thus, the Weibull family (Example 2) represents the knife-edge case for the effect of marginal regulatory interference, i.e., the effect of regulation on the second term in (27) is also zero. To provide a graphical illustration of the global effects of deferral regulation in the absence of a relevant participation constraint, the left-hand panel of Figure 4 plots the special case of the exponential distribution (Weibull with $\kappa = 1$) for $v = 0$. As

\[34\] Under an additional regularity condition (uniform convergence of the derivatives $\frac{\partial d \log L(\tau|a)}{\partial a}$), the effect is monotonic in the limit, i.e., $\lim_{\tau \to \infty} \frac{da_{RE}(\tau)}{d\tau} < 0$ (see the proof of Proposition 1).
Figure 4. Comparative statics. This graph plots equilibrium effort $a^*(\tau)$ as a function of the mandatory delay of compensation $\tau$ for the case of an exponential distribution. The left panel shows a case with low outside option, $v = 0$, such that the participation constraint is irrelevant for any $\tau$, whereas it does bind eventually in the right panel with $v = 30$. For both panels $\Delta_r = 1.25$, the revenue function satisfies $\pi(a) = 400a$ and effort costs are given by $c(a) = \frac{1}{15}a^3$.

$\tau$ starts increasing from $\tau = 0$, the optimal action choice is initially unaffected until $\tau$ reaches the unconstrained optimal payout time $T_{RE}(a|0) = \frac{1}{\Delta_r}$. Initially, the effect of binding regulation on effort is second-order, but as the effect of the timing inefficiency on the first term in (27) becomes stronger, equilibrium effort $a^*(\tau) = a_{RE}(\tau)$ is decreasing towards zero.

We will now study how the principal’s choices change in presence of a relevant participation constraint.

Relevant participation constraint. When $v > 0$, the agent’s participation constraint is relevant for sufficiently stringent regulation $\tau$. Either, (PC) is binding from the outset or, becomes relevant eventually as $\tau$ increases. In both cases, when (PC) binds, the comparative statics will crucially depend on whether the principal optimally uses a $CMI$-contract or not.

Consider, first, the case where the principal uses a $CMI$-contract. Then, for $\tau = 0$, the optimal contract implementing a given action $a$ consists of a single payment at $T_1(a)$

\[ \text{Figure 4. Comparative statics. This graph plots equilibrium effort } a^*(\tau) \text{ as a function of the mandatory delay of compensation } \tau \text{ for the case of an exponential distribution. The left panel shows a case with low outside option, } v = 0, \text{ such that the participation constraint is irrelevant for any } \tau, \text{ whereas it does bind eventually in the right panel with } v = 30. \text{ For both panels } \Delta_r = 1.25, \text{ the revenue function satisfies } \pi(a) = 400a \text{ and effort costs are given by } c(a) = \frac{1}{15}a^3. \]

$\tau$ starts increasing from $\tau = 0$, the optimal action choice is initially unaffected until $\tau$ reaches the unconstrained optimal payout time $T_{RE}(a|0) = \frac{1}{\Delta_r}$. Initially, the effect of binding regulation on effort is second-order, but as the effect of the timing inefficiency on the first term in (27) becomes stronger, equilibrium effort $a^*(\tau) = a_{RE}(\tau)$ is decreasing towards zero.

We will now study how the principal’s choices change in presence of a relevant participation constraint.

Relevant participation constraint. When $v > 0$, the agent’s participation constraint is relevant for sufficiently stringent regulation $\tau$. Either, (PC) is binding from the outset or, becomes relevant eventually as $\tau$ increases. In both cases, when (PC) binds, the comparative statics will crucially depend on whether the principal optimally uses a $CMI$-contract or not.

Consider, first, the case where the principal uses a $CMI$-contract. Then, for $\tau = 0$, the optimal contract implementing a given action $a$ consists of a single payment at $T_1(a)$

\[ \text{35 The latter result is straightforward from the fact that the optimal action choice with slack (PC) goes to zero as } \tau \text{ increases (cf. Proposition 1) such that the agent’s utility } V_A(a_{RE}(\tau), T_{RE}(a_{RE}(\tau)|\tau)) \text{ goes to zero eventually (cf. Lemma 7).} \]
solving $V_A(a, T_1(a)) = 0$, where we have used Assumption 4. It then follows directly from the monotonicity of $V_A(a, T)$ (cf. Lemma 7) that each action $a$ pins down a unique $T_1(a)$. Hence, the problem of finding the optimal action can be equivalently written as

$$\max_T \left\{ \pi(a(T)) - e^{\Delta_T} (v + c(a(T))) \right\}$$

$$\text{s.t.} \quad \frac{c'(a(T))}{L(T|a(T))} - (v + c(a(T))) = 0.$$  \hfill (29)

To facilitate exposition, we again assume that this is a strictly concave problem and denote the unique solution by $T_1^* > 0$, with associated optimal action $a(T_1^*)$. Now, for $\tau > 0$, it clearly remains optimal to set $a^*(\tau) = a(T_1^*)$ as long as $T_1^* \geq \tau$, while we have $a^*(\tau) = a(\tau)$ for any $\tau > T_1^*$, where the regulatory constraint binds. In the latter case, the optimal action along with the optimal contract are, thus, completely pinned down by the set of constraints.

**Lemma 10** For any $v > 0$, there exists a finite $\tau' \geq 0$ such that (PC) binds for all $\tau \geq \tau'$. Then, if the principal uses a $CMI$-contract and the regulatory constraint binds, the equilibrium action is given by $a(\tau)$, which is strictly increasing in $\tau$.

The intuition why the equilibrium action is increasing when the principal uses a $CMI$-contract and (PC) binds is simple: Regulation forces the principal to increase contract informativeness, while the size of the compensation package is fixed by the outside option. However, as we know from our discussion of optimal compensation design under deferral regulation, if (PC) binds, the bank might optimally adjust the contingency of payouts and, hence, deviate from $CMI$-contracts, when regulation is sufficiently stringent. More precisely, when $L(\tau|a) > \frac{c'(a)}{v + c(a)}$, the agent’s share in the firm needed to induce participation is sufficiently large such that the incentive problem is irrelevant for compensation costs (cf. Lemma 8). In this case, the optimal action choice is the solution to the following strictly concave problem:

$$\tilde{a}(\tau) = \arg \max_a \pi(a) - e^{\Delta_T} (v + c(a)).$$  \hfill (30)

Then, if $L(\tau|\tilde{a}(\tau)) \geq \frac{c'(\tilde{a}(\tau))}{v + c(\tilde{a}(\tau))}$, the incentive constraint (IC) is indeed irrelevant for compensation costs at $\tilde{a}(\tau)$ (cf. Lemma 8) and the optimal action satisfies $a^*(\tau) = \tilde{a}(\tau)$. \hfill (36)
From the monotonicity of $L$ in $\tau$, this will intuitively only be the case if regulation is stringent enough.$^{37}$

**Proposition 2** Assume $v > 0$. Then there exists a finite $\tau'' \geq \tau'$ such that the principal uses $\mathcal{C}_{M1}$-contracts if and only if $\tau \leq \tau''$. For $\tau > \tau''$ the equilibrium action satisfies $a^*(\tau) = \bar{a}(\tau)$ and is strictly decreasing in $\tau$ with $\lim_{\tau \to \infty} \bar{a}(\tau) = 0$.

That $\bar{a}(\tau)$ is decreasing in $\tau$ is obvious from the compensation cost function for the case without relevant (IC), $\tilde{W}(\tau) = \tilde{w}(\tau) (v + c(a))$, with $\tilde{w}(\tau) := e^{\Delta_r \tau}$. Similar to our discussion for the case of slack (PC) in the previous Section, the multiplicative form of $\tilde{W}(\tau)$ implies that not only the level of costs but also marginal costs of implementing a higher effort level have to increase in $\tau$. Intuitively, this is the case as discounting costs apply per unit of pay, which has to increase in the implemented action to compensate the agent for higher effort costs. As a consequence marginal compensation costs increase with $\tau$ and equilibrium effort decreases.

The right-hand panel of Figure 4 provides a graphical illustration of Proposition 2 and Lemma 10 for the case of an exponential distribution. All parameters are as in the left-hand panel but now there is a relevant participation constraint as $v = 30$. Still, (PC) is slack in the absence of regulation and the optimal payout time is again $T_{RE}(a|0) = \frac{1}{\Delta_r}$. When regulation starts to bind, $a^*(\tau) = a_{RE}(\tau)$ decreases and so does the agent’s value $V_A(a^*(\tau), \tau)$ until the participation constraint binds which occurs at $\tau = \tau'$, where $v_{RE}(\tau') = V_A(a_{RE}(\tau'), \tau')$ is exactly equal to $v$. As $\tau$ increases further, both (PC) and (IC) are relevant for compensation costs and, together with (R), pin down the optimal action choice completely $a^*(\tau) = a(\tau)$, which from (29) is increasing in $\tau$ (cf. Lemma 10). However, eventually, at $\tau = \tau''$, the regulatory constraint increases contract informativeness to a level where (IC) becomes irrelevant for compensation costs, given the principal has to satisfy (PC). Put differently, for any $\tau > \tau''$, the principal deviates from $\mathcal{C}_{M1}$-contracts and his optimal action choice maximizes efficiency subject to the regulatory constraint, i.e., $a^*(\tau) = \bar{a}(\tau)$, which from (30) is strictly decreasing in $\tau$ (cf. Proposition 2).

**Lessons for regulation.** The results obtained so far are no good news for deferral regulation. In fact, without precise knowledge of the arrival time distribution, marginal regulation has an ambiguous effect on the equilibrium action, which might change non-monotonically as regulation becomes more stringent. Further, combining Propositions 1 and 2 we get the following robust result:

$^{37}$E.g. in the boundary case where $\tau \to 0$, it follows from $\bar{a}(\tau) \to a^{FB}$, where $a^{FB} = \arg\max_a \{ \pi(a) - (v + c(a)) \}$, together with $L(0|a^{FB}) = 0$ that $v(\tau) \to \infty$.  

33
Corollary 2 For sufficiently stringent regulation, the equilibrium action satisfies, for any value $v$ of the agent’s outside option, $\lim_{\tau \to \infty} a^*(\tau) = 0$.

I.e., sufficiently stringent regulation unambiguously decreases the equilibrium action towards zero, such that, taken together, pure deferral regulation aiming at increasing the equilibrium action is likely to backfire.

However, deferral regulation is not a complete lost cause. As we will show next, complementing deferral regulation with a restriction on the contingency of pay is more likely to be successful in increasing the equilibrium action. So, consider the following additional regulatory constraint, requiring the absence of a disaster at the stipulated payout time $t$, i.e.,

$$db(t) = 0 \quad \forall t \geq \tau \text{ if } Y(t) = 1.$$  \hspace{1cm} (CLAW)

We interpret this additional constraint as a clawback clause, as one may implement the contractual restrictions via an escrow account subject to clawback provisions.

One naively might expect that, when forced to defer compensation, the principal will optimally make use of the additional information to provide stronger incentives, such that the additional clawback clause does not bind under the optimal contract. And indeed, as long as the equilibrium action is implemented with a $CMI$-contract, (CLAW) plays no role in our analysis. The principal endogenously makes all pay conditional on survival anyway. However, from Proposition 2 the principal deviates from $CMI$-contracts and makes a fraction of pay unconditional if $v > 0$ and regulation becomes sufficiently stringent. In that case the principal does not fully use the forced increase in informativeness to provide stronger incentives, but rather implements a lower effort $\tilde{a}(\tau) < a(\tau)$, in order to reduce effort costs for which the agent has to be compensated. It is precisely in this case where (CLAW) ”bites.” By requiring to use $CMI$-contracts, i.e., to make all pay contingent on the most informative history, the principal is “forced” by regulation to increase his action as it is no longer determined by a first-order condition, but instead given by the lowest possible implementable action that allows satisfying (PC), (IC), (R) and (CLAW), i.e., $a^*(\tau) = a(\tau)$, which is strictly increasing in $\tau$ (cf. Lemma 10).

To understand the intuition, recall that regulatory interference in the timing of pay does not give the principal per se more incentives to avoid disaster events (in contrast to, for example, equity capital regulation). In particular, when facing regulatory constraints on one dimension of the compensation contract, the principal will, in general, try to adjust other dimensions of the contract. In our setup, the relevant dimensions are 1) the

\footnote{That is, the principal makes pay contributions into an escrow account even before date $\tau$ (yielding interest at rate $r_P$). However, the agent only has unrestricted access to the account after date $\tau$ and if the institution has not failed at the time of the stipulated payout.}
size of the agent’s compensation package $B$ and 2) the contingency of pay (success versus failure). If $\tau$ is sufficiently high, the principal now is unable to adjust on any margin. First, binding (PC) fixes $B$ at $v + c(a^*(\tau))$. Second, (CLAW) forces the principal to only pay in the absence of the disaster, which combined with the deferral requirement $\tau$, makes it impossible to implement a sufficiently low action. Taken together, in this case the principal has to make a compensation package of fixed size contingent on a more informative signal, which mechanically leads to higher equilibrium effort. To see this graphically, reconsider the right-hand panel of Figure 2. While, without a clawback clause the principal optimally deviates from maximum incentives contracts for $\tau > \tau''$, causing $a^*(\tau)$ to decrease eventually, imposing (CLAW) forces him to implement $a(\tau)$ for any $\tau > \tau''$, such that sufficiently stringent regulation increases the equilibrium action.

4 Conclusion

tbc
Appendix A  Proofs

Proof of Lemma A1. The proof proceeds in two steps: First, we show, that it is never optimal to make strictly positive payments following any history \( h^t \neq h_{MI}^t \) for \( t > 0 \) (cf. Lemma A1 below). Second, we show that strictly positive payments following \( h^0 \neq h_{MI}^0 \) are never optimal if action \( a \) can be implemented with a \( \mathcal{C}_{MI} \)-contract (cf. Lemma A2 below).

**Lemma A1** The optimal contract provides maximal incentives for any \( t > 0 \), i.e., it holds that \( db(t|h^t) = 0 \) for all \( h^t \neq h_{MI}^t, t > 0 \).

Proof of Lemma A1. The proof is by contradiction. So, assume to the contrary that the optimal contract has for some \( t > 0, db(t|h^t) > 0 \) with \( h^t \neq h_{MI}^t \). We distinguish two cases, depending on whether \( \frac{\partial \Pr(h^t|a)}{\partial a} \geq 0 \). Assume, first, that \( \frac{\partial \Pr(h^t|a)}{\partial a} \geq 0 \). Then, there exists a feasible perturbation of the initial contract with lower expected compensation costs to the principal. To construct such a perturbation, decrease \( db(t|h^t) \) to zero, increase \( db(t|h_{MI}^t) \) by \( \Delta db(t|h_{MI}^t) \) to satisfy (IC), and make an additional unconditional payment at \( t = 0 (\forall h^0 \in H^0) \) of \( \Delta b_0 \) to keep \( V_A \) constant and satisfy (PC). Thus, we require

\[
\Delta b_0 + e^{-r_A t} \left[ \Pr(h_{MI}^t|a) \Delta db(t|h_{MI}^t) - \Pr(h^t|a) db(t|h^t) \right] = 0,
\]

\[
\frac{\partial \Pr(h_{MI}^t|a)}{\partial a} \Delta db(t|h_{MI}^t) - \frac{\partial \Pr(h^t|a)}{\partial a} db(t|h^t) = 0,
\]

Solving for \( \Delta b_0 \) and \( \Delta db(t|h_{MI}^t) \) reveals that both adjustments are positive, and hence do not violate (LL). The resulting impact on wage cost is:

\[
\Delta W = (e^{-r_P t} - e^{-r_A t}) \Pr(h^t|a) \left[ \frac{\partial \Pr(h^t|a)}{\partial a} / \Pr(h^t|a) - 1 \right] db(t|h^t) < 0,
\]

where we have used (6) as well as \( \Delta r = r_A - r_P > 0 \), contradicting optimality of the initial contract. The proof for the remaining case where \( \frac{\partial \Pr(h^t|a)}{\partial a} < 0 \) is immediate, as in this case setting \( db(t|h^t) = 0 \) relaxes the incentive constraint, thus allowing to reduce any incentive pay while ensuring that the participation constraint continues to hold by making an additional unconditional payment at time 0, which reduces costs resulting from impatience. We omit a formal argument for brevity. ■

So far we have shown that any contract that pays after history \( h^t \neq h_{MI}^t \) for date \( t > 0 \) is dominated. We will next show that, if this does not violate implementability of
a, we also have $db(0|h^0) = 0$ for all $h^0 \neq h^0_{MI}$ under the optimal contract. To do so, it is useful to define $L(t|a) := \left[ \frac{\partial}{\partial a} \Pr (h^t_{MI}|a) \right] / \Pr (h^t_{MI}|a)$ as the maximum likelihood ratio for each given $t$. The result then trivially holds whenever there is no information arriving at $t = 0$, i.e., if $L(0|a) = 0$, such that $h^0 = h^0_{MI}$ for all $h^0 \in H^0$. In the following we will, hence, restrict attention to the case where $L(0|a) > 0$. Then, we have the following result:

**Lemma A2** If $a$ satisfies $v + c(a) - \frac{c'(a)}{L(0|a)} \leq 0$, then $db(0|h^0) = 0$ for all $h^0 \neq h^0_{MI}$.

**Proof of Lemma A2.** The proof is by contradiction. So, assume to the contrary that the optimal contract has $db(t|h^0) > 0$ for some $h^0 \neq h^0_{MI}$. We again focus on the case with $\frac{\partial Pr\left(h^0_t|a\right)}{\partial a} \geq 0$ (the case with $\frac{\partial Pr\left(h^0_t|a\right)}{\partial a} \leq 0$ is a straightforward extension). Then there exists a sequence of feasible perturbations of the initial contract resulting in lower expected compensation costs to the principal. To see this, assume, first, that (PC) is slack under the candidate contract. Then reduce $db(0|h^0)$ by $\Delta db(0|h^0)$ and increase $db(0|h^0_{MI})$ by $\Delta db(0|h^0_{MI}) = \left( \frac{\partial Pr\left(h^0_t|a\right)}{\partial a} / \frac{\partial Pr\left(h^0_{MI}|a\right)}{\partial a} \right) | db(0|h^0) >= 0$, which ensures that (IC) is satisfied. The resulting change in wage costs is given by

$$\Delta W = Pr\left(h^0_t|a\right) \left[ \frac{\partial Pr\left(h^0_t|a\right)}{\partial a} / L(0|a) \right] \Delta db(0|h^0) < 0,$$

where we have used (0), contradicting optimality of the initial contract. If one can pick $\Delta db(0|h^0) = db(0|h^0)$ and (PC) is still satisfied the result follows. Else, pick $\Delta db(0|h^0)$ such that (PC) holds with equality following the perturbation and proceed as follows.

When $db(0|h^0) > 0$ and (PC) binds, this implies from the parameter restriction $v + c(a) - \frac{c'(a)}{L(0|a)} \leq 0$, that there must be at least one strictly positive bonus payment $db(t|h^t_{MI})$ for some $t > 0$ with $L(t|a) > L(0|a)$.

Denote the set of dates $t > 0$ with strictly positive bonus payments by $T > 0$ and pick some $t_1 \in T > 0$. Then consider the

$$\sum_{h^0 \in H^0 \backslash h^0_{MI}} \Pr\left(h^0_t|a\right) db(0|h^0) + \int_{[0,T]} e^{-\gamma T} \Pr\left(h^t_{MI}|a\right) db\left(t|h^t_{MI}\right) = v + c(a),$$

$$\sum_{h^0 \in H^0 \backslash h^0_{MI}} \frac{\partial Pr\left(h^0_t|a\right)}{\partial a} db(0|h^0) + \int_{[0,T]} e^{-\gamma T} \frac{\partial Pr\left(h^t_{MI}|a\right)}{\partial a} db\left(t|h^t_{MI}\right) = c'(a).$$

Dividing the second equation by $L(0|a) > 0$ and subtracting from the first gives, after some algebra,

$$\sum_{h^0 \in H^0 \backslash h^0_{MI}} \Pr\left(h^0_t|a\right) \alpha(h^0) db(0|h^0) + \int_{[0,T]} e^{-\gamma T} \Pr\left(h^t_{MI}|a\right) \beta(t) db\left(t|h^t_{MI}\right) = v + c(a) - \frac{c'(a)}{L(0|a)}, \ (31)$$

39 To see this, write the binding (PC) and (IC) constraints as

$$\sum_{h^0 \in H^0 \backslash h^0_{MI}} \Pr\left(h^0_t|a\right) db(0|h^0) + \int_{[0,T]} e^{-\gamma T} \Pr\left(h^t_{MI}|a\right) db\left(t|h^t_{MI}\right) = v + c(a),$$

$$\sum_{h^0 \in H^0 \backslash h^0_{MI}} \frac{\partial Pr\left(h^0_t|a\right)}{\partial a} db(0|h^0) + \int_{[0,T]} e^{-\gamma T} \frac{\partial Pr\left(h^t_{MI}|a\right)}{\partial a} db\left(t|h^t_{MI}\right) = c'(a).$$

Dividing the second equation by $L(0|a) > 0$ and subtracting from the first gives, after some algebra,

$$\sum_{h^0 \in H^0 \backslash h^0_{MI}} \Pr\left(h^0_t|a\right) \alpha(h^0) db(0|h^0) + \int_{[0,T]} e^{-\gamma T} \Pr\left(h^t_{MI}|a\right) \beta(t) db\left(t|h^t_{MI}\right) = v + c(a) - \frac{c'(a)}{L(0|a)}, \ (31)$$
following perturbation of the candidate contract: Reduce \( db(0|h_0^0) \) by \( \Delta db(0|h_0^0) \) and the bonus payment at \( t_1 > 0 \), by \( \Delta db(t_1|h_{MI}^{t_1}) \), while increasing \( db(0|h_{MI}^0) \) by \( \Delta db(0|h_{MI}^0) \). In particular, in order to continue satisfying (PC) and (IC) we require

\[
\Pr (h_{MI}^0 | a) \Delta db(0|h_{MI}^0) - \Pr (\tilde{h}_0 | a) \Delta db(0|\tilde{h}_0) - e^{-r_{f_1}} \Pr (h_{MI}^{t_1} | a) \Delta db(t_1|h_{MI}^{t_1}) = 0,
\]

which follows from the agent’s relative impatience, contradicting optimality of the initial contract. If \( \Delta db(0|\tilde{h}_0) = db(0|\tilde{h}_0) \) implies from (32) that \( \Delta db(t_1|h_{MI}^{t_1}) \leq db(t_1|h_{MI}^{t_1}) \) the result follows. Else, pick \( \Delta db(0|\tilde{h}_0) \) such that \( \Delta db(t_1|h_{MI}^{t_1}) = db(t_1|h_{MI}^{t_1}) \). The resulting contract (still) has \( db(0|\tilde{h}_0) > 0 \) and there thus exists \( t_2 \in T^{>0}\setminus t_1 \). The result then follows by iterating the same steps as above until \( db(0|\tilde{h}_0) = 0 \).

It remains to show that \( a \) is not implementable with a \( \mathcal{C}_{MI} \)-contract if \( v + c(a) - \frac{c'(a)}{L_0(a)} > 0 \). This however follows directly from monotonicity of \( L(t|a) \). This completes the proof. Q.E.D.

with \( \alpha(h_0) = 1 - \frac{\partial \Pr (h_0 | a)}{\partial a} \frac{L_0(a)}{L(t|a)} > 0 \) for all \( h_0 \in H_0 \setminus h_{MI}^0 \) and \( \beta(t) = 1 - \frac{L(t|a)}{L(t|a)} \leq 0 \) for \( t > 0 \), where the inequalities follow from (6) and the definition of \( L(t|a) \). Now as the right hand side is non-positive by assumption, and \( db(0|\tilde{h}_0) > 0 \), we must have \( db(t|h_{MI}^t) > 0 \) for some \( t \) where \( \beta(t) > 0 \). If no such \( t \) exists, i.e., if no information arrives after \( t = 0 \), then the model is static and the result follows from standard arguments.

\[ 40 \text{Note that under a } \mathcal{C}_{MI} \text{-contract (PC) and (IC) are given by} \]

\[
\begin{align*}
\int_{[0,T]} e^{-r t} \Pr (h_{MI}^t | a) \, db (t|h_{MI}^t) &\geq v + c(a), \\
\int_{[0,T]} e^{-r t} \frac{\partial \Pr (h_{MI}^t | a)}{\partial a} \, db (t|h_{MI}^t) &\geq c'(a).
\end{align*}
\]
Proof of Lemma 2. Denote by \( h^t \) the history at \( t \) before time-\( t \) uncertainty has realized, i.e., \( h^t = \lim_{s \uparrow t} h^s \). Then, from our assumption that \( \Pr (h^t | a) > 0 \) for all \( h^t \) and \( a \), we also must have that \( \Pr (x_t | h^t ; a) > 0 \) for all possible realizations \( x_t \in X^t \) of the underlying signal process at \( t \). Further, these conditional probabilities have to satisfy for all \( a \) that \( \sum_{x_t \in X^t} \Pr (x_t | h^t ; a) = 1 \), implying

\[
\sum_{x_t \in X^t} \frac{\partial \Pr (x_t | h^t ; a)}{\partial a} = 0.
\]

So, either \( \frac{\partial \Pr (x_t | h^t ; a)}{\partial a} = 0 \) for all \( x_t \) or, \( \exists x_t \in X^t \) satisfying \( \frac{\partial \Pr (x_t | h^t ; a)}{\partial a} > 0 \). Now, from \( \Pr (h^t | a) = \Pr (h^t | a) \Pr (x_t | h^t ; a) \), we obtain

\[
\frac{\partial \Pr (h^t | a)}{\partial a} = \frac{\partial \Pr (h^t | a)}{\partial a} + \frac{\partial \Pr (x_t | h^t ; a)}{\partial a}.
\]

Setting \( h^{-} = h^{-}_{MI} \), there, thus, exists a \( x_t \in X^t \) such that

\[
\frac{\partial \Pr (h^t | a)}{\partial a} = L(t^{-} | a) + \frac{\partial \Pr (x_t | h^t ; a)}{\partial a} \geq 0,
\]

i.e., there exist a (continuation) history \( h^t \) with a likelihood ratio (weakly) greater than \( L(t^{-} | a) \). Q.E.D.

Proof of Lemma 3. See main text. Q.E.D.

Proof of Theorem 1. That \( a \) is implementable with a \( CI_M \)-contract if and only if \( L(0 | a) \leq \frac{c(a)}{v + c(a)} \) follows from the same arguments as in the proof of Lemma 1. When \( L(0 | a) \leq \frac{c(a)}{v + c(a)} \), Lemma 1 then implies that the optimal contract must be a \( CI_M \)-contract and hence solves Problem 1*. Consider, 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 (9) which implies from (IC) that \( B = c'(a) / L(T_{RE}(a) | a) \). 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 \( v \leq \bar{v} := c'(a) / L(T_{RE}(a) | a) - c(a) \). Else, it is easy to show that \( PC \) must bind under the optimal contract, i.e., \( B = v + c(a) \) and the

\[
\int_{0}^{T} e^{-r t} \Pr (h^t | a) \left[ 1 - \frac{L(t | a)}{L(0 | a)} \right] db(t | h^t_{MI}) \geq v + c(a) - \frac{c'(a)}{L(0 | a)}
\]

where the left-hand side is negative from (LL) together with the monotonicity of \( L(t | a) \).
optimal payout times are as characterized in Lemma 3. Q.E.D.

**Proof of Lemma 4.** That $a$ is not implementable with a $C_M$-contract when $L(0|a) > \frac{c'(a)}{v + c(a)}$ follows from the arguments in the proof of Lemma 4. Note next that, due to differential discounting and $(PC)$, there is a lower bound on wage costs given by $W(a) = B = v + c(a)$. Now, consider the contract specified in the main text, which consists of a single bonus payment at time 0 of $db(0|h_{M1}) = \gamma_{M1}B/\Pr(h_{M1}|a)$, where $\gamma_{M1}$ is given in (12), next to an unconditional payment at time 0 (for all possible histories $H_0$) of $b_0 = v + c(a) - \frac{c'(a)}{L(0|a)} > 0$, where the inequality follows by assumption. This contract satisfies $(PC)$, $(IC)$ and $(LL)$ and implies wage costs of $W(a) = W'(a) = v + c(a)$. Hence, it solves the compensation design problem. The result then follows by noting that any contract with strictly positive payments at $t > 0$ must necessarily imply higher wage costs due to relative impatience. Q.E.D.

**Proof of Lemma 5.** To streamline the derivations in the remainder of this appendix, let a subscript denote partial derivatives, e.g., $S_a(t|a) := \frac{\partial S(t|a)}{\partial a}$. Direct differentiation of (20) gives (21) and monotonicity of $L(t|a)$ follows from Assumption 3. Then it remains to show that the $h_{M1}'$ history is indeed “absence of failure.” So consider any alternative history $h' \in H$, which corresponds to failure at some $t' \leq t$ and, thus, has a likelihood ratio of $f_a(t'|a)/f(t'|a)$. Now, Assumption 3 implies that

$$\frac{\partial}{\partial a} f(t'|a) S(t'|a) = \frac{f(t'|a)}{S(t'|a)} \left( \frac{f_a(t'|a)}{f(t'|a)} - L(t'|a) \right) < 0,$$

and the result follows from monotonicity of $L(t|a)$. Q.E.D.

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

**Proof of Lemma 7.** The comparative statics in $T$ follow directly from strict monotonicity of $L(T|a)$ (cf. Lemma 5). For the marginal effect of $a$ note that we obtain after some algebra

$$\frac{\partial V_A(a, T)}{\partial a} = \frac{c'(a)}{L(T|a)} \left( \frac{c''(a)}{c'(a)} - \frac{S_{aa}(T|a)}{S_a(T|a)} \right).$$

(33)

Now note that (local) incentive compatibility of a $C_M$-contract implementing effort $a$ with a single payout of $b$ at time $T$ requires that both the first-order condition

$$\left. \frac{\partial V_A(\hat{a}, T)}{\partial \hat{a}} \right|_{\hat{a} = a} = e^{rA}S_a(T|a)b - c'(a) = 0,$$

(34)
as well as the second-order condition of the agent’s maximization problem

\[
\frac{\partial^2 V_A(\tilde{a}, T)}{\partial \tilde{a}^2} \bigg|_{\tilde{a} = a} = e^{a^T S a} b - c''(a) \leq 0,
\]

(35)

are satisfied. Substituting \(34\) into \(35\) we obtain

\[
c'(a) \left( \frac{S_{aa}(T|a)}{S_a(T|a)} - \frac{c''(a)}{c'(a)} \right) \leq 0
\]

(36)

such that the expression in \(33\) has to be non-negative. Q.E.D.

**Proof of Lemma 8.** See main text and the proof of Lemma 4. Q.E.D.

**Proof of Lemma 9.** When \(v \leq v_{RE}(\tau)\), then \(\text{(PC)}\) is satisfied at the solution \(a_{RE}(\tau)\) to the relaxed problem (ignoring the participation constraint), such that the relaxed and the full problem are in fact equivalent. It remains to show, that in all other cases the participation constraint is binding. This follows, however, directly from strict concavity of the relaxed problem. Q.E.D.

**Proof of Proposition 1.** By the implicit function theorem,

\[\text{sgn} \left( \frac{da_{RE}(\tau)}{d\tau} \right) = \text{sgn} \left( -\frac{\partial^2 W_{RE}(a|\tau)}{\partial a \partial \tau} \right),\]

with the cross-partial given by

\[
\frac{\partial^2 W_{RE}(a|\tau)}{\partial a \partial \tau} = 1_{T_{RE}(a|\tau) = \tau} \left[ \left( \Delta_r - \frac{d \log L}{dt} \bigg|_{t=\tau} \right) \left[ \frac{c''(a)}{c'(a)} - \frac{L_a}{L} \right] - \frac{\partial}{\partial a} \frac{d \log L}{dt} \bigg|_{t=\tau} \right].
\]

(37)

The expression in \(28\) for \(\tau = T_{RE}(a|0)\) then follows from \(10\) and the ambiguous effects of marginal regulation are illustrated e.g. in the following Corollary. It remains to show that \(\lim_{\tau \to \infty} a_{RE}(\tau) = 0\). To see this rewrite \(27\) to obtain

\[
\frac{\partial W_{RE}(a|\tau)}{\partial a} = \left[ \frac{1}{L} \left( \frac{c''(a)}{c'(a)} - \frac{S_{aa}}{S_a} \right) + 1 \right] e^{\Delta_r T_{RE}(a|\tau)} c'(a),
\]

where, from \(L(T_{RE}(a|\tau)|a) > 0\) and local incentive compatibility, \(36\), the term in square brackets is strictly positive. Hence, \(T_{RE}(a|\tau) \geq \tau\) directly implies that marginal costs go to infinity as \(\tau \to \infty\) for any \(a > 0\), and the result follows from \(25\) together with the assumptions on \(\pi(a)\). Q.E.D.
Proof of Lemma 10. From Lemma 9, (PC) binds for all $v > v_{RE}(\tau)$. Existence of a finite threshold $\tau'$ then follows from $\lim_{\tau \to \infty} v_{RE}(\tau) = \lim_{\tau \to \infty} V_A(a_{RE}(\tau), \tau) = 0$, where we have used $\lim_{\tau \to \infty} a_{RE}(\tau) = 0$ together with Lemma 7. It remains to show monotonicity of $a(\tau)$ for $T > 0$. This follows directly from $\frac{\partial a(T)}{\partial \tau} = L_T / \left( T \left( L \left( \frac{\partial^2 a(T)}{\partial \tau^2} - \frac{S_{IM}}{S_{MI}} \right) \right) \right)$ together with agent optimality (cf. (36)). Q.E.D.

Proof of Proposition 2. It is obvious from (30) that $\tilde{a}(\tau)$ is strictly decreasing in $\tau$ and approaches zero for $\tau \to \infty$. Now note that $\tilde{a}(\tau)$, the action choice that is optimal in the relaxed problem without (IC), can be implemented at same wage costs if and only if $L(\tau|\tilde{a}(\tau)) \geq \frac{c'(\tilde{a}(\tau))}{v + c(\tilde{a}(\tau))}$ which is equivalent to $v \geq \bar{v}(\tau) := V_A(\tilde{a}(\tau), \tau)$. Existence of a finite threshold $\tau''$ then follows from $\lim_{\tau \to \infty} \bar{v}(\tau) = 0$, where we have used $\lim_{\tau \to \infty} \tilde{a}(\tau) = 0$ together with Lemma 7 and recalling the results on the optimality of $C_{MI}$-contracts in Lemmas 1 and 4. It remains to show that the thresholds satisfy $\tau' \leq \tau''$. To see this, denote the objective function in (24) by $\Pi_{RE}(a)$, which can be rewritten to obtain

$$\Pi_{RE}(a) = \pi(a) - e^{\Delta^{T_{RE}(a|\tau)}(v + c(a))} - e^{\Delta^{T_{RE}(a|\tau)}} [V_A(T_{RE}(a|\tau)|a) - v].$$

Differentiating with respect to $a$ we get

$$\Pi'_{RE}(a) = \pi'(a) - e^{\Delta^{T_{RE}(a|\tau)}c'(a)} - e^{\Delta^{T_{RE}(a|\tau)}} \frac{\partial V_A(T_{RE}(a|\tau)|a)}{\partial a},$$

where we have used the envelope theorem. Now, as the agent’s optimal action choice (FOA) implies that $\frac{\partial V_A(T_{RE}(a|\tau)|a)}{\partial a} \geq 0$ (cf. Lemma 7), it follows that $a_{RE}(\tau) = \arg\max_a \Pi_{RE}(a) \leq \arg\max_a \{ \pi(a) - e^{\Delta^{T_{RE}(a_{RE}(\tau)|\tau)}} (v + c(a)) \} \leq \arg\max_a \{ \pi(a) - e^{\Delta^{T}} (v + c(a)) \} = \tilde{a}(\tau)$, where, for the second inequality, we have used $T_{RE}(a_{RE}(\tau)|\tau) \geq \tau$. Finally, note that the inequality is strict, and hence $a_{RE}(\tau) < \tilde{a}(\tau)$, if $T_{RE}(a_{RE}(\tau)|\tau) > \tau$, which is clearly satisfied at $\tau = 0$. It then follows from Lemma 7 that $\bar{v}(\tau) \geq v_{RE}(\tau)$ or equivalently $\tau' \leq \tau''$. Q.E.D.

Appendix B Supplementary Material

B.1 Multitask Environment

Consider the following multi-task environment inspired by Bénabou and Tirole (2016). The agent’s action consists of two tasks $a$ and $q$, where only task $a$ has persistent effects. 

\footnote{41 Clearly the regulatory constraint can only bind for $\tau > 0$.}
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

\[
W(a,q|\tau) = qB \min_{B,w(t)} \int_0^\infty e^{\Delta t} dw(t) \quad \text{s.t.}
\]

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

\[
B \int_0^\infty L(t|a) dw(t) = \frac{\partial c(a,q)}{\partial a} \quad \text{(ICa)}
\]

\[
dw(t) \geq 0 \quad \forall t \quad \text{(LL)}
\]

Since \( a \) and \( q \) are fixed, it is immediate that the solution to this compensation design problem is given by Lemma 3 where ICq has similar effects as a binding participation constraint.

### B.2 Application - Example Distributions

In this Appendix we provide some additional material for our main application to the financial sector in Section 3 specifying concrete arrival time distributions.

**Hazard rates, likelihood ratios and effort.** Figure 5 plots the hazard rate and function \( L \) for specifications of Example distributions 1 to 3. We plot the functions for two effort levels. It can be inferred from the graph that the slope of \( L \) approaches zero the smaller the difference in hazard rates under high and low effort.

**Optimal payout times.** In the absence of regulation, Theorem applies one-to-one. Since \( L(t|\cdot) \) is differentiable in this application, the payout time with slack (IP), \( T_{RE}(a) \),
is characterized by the first-order condition in (10). For the Weibull family considered in Example 2, this trade-off yields a particularly simple closed-form solution $T_{RE}(a) = \frac{\kappa}{\Delta_r}$, which nests the exponential distribution for $\kappa = 1$. For the mixed exponential distribution, which is a concrete case of the more general mixed distribution in Example 1, we similarly obtain $T_{RE}(a) = \frac{1}{\Delta_{\lambda}} \ln \left( 1 + \frac{\Delta_{\lambda} - \Delta_r + \sqrt{(\Delta_{\lambda} - \Delta_r)^2 + 4\Delta_{\lambda} \Delta_r}}{2\Delta_r} \right)$, where $\Delta_{\lambda} := \lambda_H - \lambda_L$.

When (PC) binds, the maximum number of payments is generally two (see Theorem 1). Since informativeness is continuous in our application (see Figure 5) with $L(T_1|a) = 0$, there always exists a unique payout time $T_1(a)$ solving $L(T_1|a) = \frac{c'(a)}{v+c(a)}$ if (PC) binds.\(^{43}\) Whether this single payout date is optimal depends on how the growth

\(^{42}\)Concretely, in this case we have $S_L(t) = e^{-\lambda_L t}$ and $S_H(t) = e^{-\lambda_H t}$, with $\lambda_H > \lambda_L$.

\(^{43}\)Note, if $\lim_{t \to \infty} L(T|a) < \frac{c'(a)}{v+c(a)}$, then (PC) must be slack for action $a$. 

---

**Figure 5. Example information processes.** This graph plots the hazard rate (top panels) and likelihood ratio (lower panels) for different arrival time distributions falling in the class of Examples 1 to 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.
rate of informativeness changes with time. As can be inferred from Figure 5, the exponential distribution has the special feature that informativeness grows linearly. As a result, $C(L|\cdot)$ 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 the combination 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 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 $\frac{1-e^{\Delta \lambda T}}{1-e^{\Delta \lambda T_1}} = \frac{\Delta \lambda}{\Delta \lambda}$ are optimal if $a < \frac{1}{2} \left(1 - \frac{\Delta \lambda}{\Delta \lambda}\right)$, while if the latter condition is violated or the outside option is still sufficiently low, $v_1 < v_2$, there is a single optimal payout date $T_1(a) = \frac{1}{\Delta \lambda} \ln \left(1 + \frac{v'_1(a) + c'(a)}{v+ c(a) - c'(a) a}\right)$.

While we do not obtain closed form solutions for the lognormal distribution in Example 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.

**Further example distributions.**

**Example 4** Piecewise mixed exponential: The arrival time distribution satisfies $F(t|a) = wG(t|a)$ for $t \leq s$ and $F(t|a) = F(s|a) + wG(t-s|a)$ 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 6 plots the information process corresponding to the piecewise mixed exponential arrival time distribution in Example 1, which features two phases of high growth. Hence, the cost of informativeness plotted in the right panel exhibits a non-convex region. In this example, the optimal unconstrained payout time (pinned down by the tangent line through the origin with the minimum slope, $\frac{e^{\Delta \lambda T}}{E(t|a)}$) 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.

---

44 If the outside option is below $v_1$, $T_{PC}$ is slack. The two thresholds on $v$ satisfy $v_1 = c'(a) \left(a + \frac{1}{e^{\Delta \lambda T_{RE}(a|0)}-1}\right) - c(a)$ and $v_2 = c'(a) \left(a + \frac{1}{e^{\Delta \lambda T(\tau-a)}-1}\right) - c(a)$, for $a < \frac{1}{2} \left(1 - \frac{\Delta \lambda}{\Delta \lambda}\right)$. 

45
Figure 6. Payout times under deferral regulation. This graph plots the likelihood ratio and the cost of delay as a function of time (left panel) as well the cost of informativeness (right panel) for the piecewise mixed exponential arrival time distribution from Example 4. The parameter values are set to $\lambda_L = 1/4$, $\lambda_H = 4$, $s = 0.71$, $a = 0.95$ and $\Delta_r = 1.55$.

References


