Regulating Deferred Incentive Pay*

Florian Hoffmann† Roman Inderst‡ Marcus Opp§

June 2014

Abstract

Our paper evaluates recent regulatory proposals mandating the deferral of bonus payments and claw-back clauses in the financial sector. We study a broadly applicable principal agent setting, in which the agent exerts effort for an immediately observable task (acquisition) and a task for which information is only gradually available over time (diligence). Optimal compensation contracts trade off the cost and benefit of delay resulting from agent impatience and the informational gain. Mandatory deferral may increase or decrease equilibrium diligence depending on the importance of the acquisition task. In particular, higher acquisition incentives make it more likely that deferral regulation increases equilibrium diligence. Our results imply concrete conditions on economic primitives that make mandatory deferral socially (un)desirable.

Keywords: Financial regulation, compensation design, principal-agent models.

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

*We are grateful to Jeff Zwiebel for very helpful suggestions on an earlier draft of the paper. In addition, we thank seminar participants at the University of Amsterdam, Berkeley, UPF, the 2013 OxFIN conference, the 2013 INF/Stockholm conference on Finance and IO, and Stanford. Hoffmann and Inderst gratefully acknowledge financial support from the ERC (Advanced Grant "Regulating Retail Finance").

†Johann Wolfgang Goethe University Frankfurt. E-mail: fhoffmann@finance.uni-frankfurt.de.

‡Johann Wolfgang Goethe University Frankfurt and Imperial College London. E-mail: inderst@finance.uni-frankfurt.de.

§Haas School of Business, University of California, Berkeley. E-mail: mopp@haas.berkeley.edu.

1 Introduction

Compensation in the financial industry has come under intense regulatory scrutiny. In particular, short-term-oriented bonus payments and commissions are blamed to have contributed both to excessive risk taking in the industry and to egregious cases of misselling of financial products to households. A key regulatory proposal is thus to mandate back-loading of compensation. Thereby, contingent compensation remains longer "at risk" in case of serious future underperformance such as insolvency of the institution or, at the retail end, default or cancellation of individual products such as mortgages, life insurances, or pension plans.

Our contribution speaks to this proposal, as we show when mandating deferred incentive pay is likely to increase the diligence with which agents conduct their business and when, instead, such regulation will backfire and decrease diligence in equilibrium. We use a model of compensation design that combines three key elements that seem important to address these issues. First, we allow the bank to compensate the respective agent at any point in time, conditional on all available performance-relevant information. Second, the bank must incentivize both the acquisition of deals or growth opportunities, as well as the exercise of diligence. Depending on the application, diligence can be directed to the choice of business strategy or to the provision of good advice and the screening of risky deals or borrowers. Third, diligence reduces the likelihood with which a (possibly rare) negative event occurs that involves a critical loss either for customers or society as a whole and that the bank and its agent do not sufficiently internalize. The last feature generates scope for regulatory interference in the first place.

The topicality of our analysis is evident from numerous regulatory initiatives around the world, all targeted towards changing the structure of compensation in the financial industry. At the level of executive pay, many reports have asserted that current compensation practices in banking are flawed and have thus proposed mandatory deferral of bonuses or mandating clawback clauses.[1] Since the G20 in Pittsburgh endorsed the FSB principles for sound compensation practices in September 2009, several policies have already been adopted. In the EU, a new directive adopted in 2010 includes strict rules

[1]See, for instance, the Squam Lake Working Group's 2010 report on Financial Regulation or for a comprehensive list of proposals, the Financial Stability Board's thematic review on compensation in their 2011 Peer Review Report.

for bank executives' bonuses.[2] At the retail end of the financial industry, some countries such as the UK have moved towards banning commissions[3], while other authorities have taken less drastic steps aimed at altering the balance of incentives through reducing the prevalence of initial commissions.[4] Even outside the financial industry, there is an ongoing debate about whether the present design of executive pay, including its timing, reflects firms' and society's interests.[5] 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.

One of the agent's tasks in our model, next to that of generating deals or acquiring customers, is to exert (more) diligence so as to make the occurrence of a possibly rare but observable (and for the bank, its customers, or society critical) event less likely. This could be the insolvency of the whole institution, the default of an individual loan, or the cancellation of a pension or insurance contract after the customer found it to be unsuitable for his needs. The optimal compensation must address both the acquisition as well as the diligence task, and it must do so while trading-off the benefits and costs of deferred compensation: More information but higher costs of delay, as we assume the agent to be more impatient than the bank. We analyze the determinants of the optimal timing of (long-term) compensation and of the weights that are given to up-front versus long-term bonus pay. Our analysis is then put to deriving the implications of mandating a longer deferral of contingent pay. Under such regulation, a bonus must be postponed until a stipulated minimum time or it must be made with the provision that it can still be clawed back until then. By imposing a minimum period until which such a bonus truly "vests," this regulation ensures that more information about the quality of the respective deal or the business as a whole comes to light before compensation is paid out. Surprisingly, we find that this does not necessarily lead to higher diligence. The reason

[2] 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. Though there are national differences, it has lead to long deferral and retention requirements (e.g., 5 years in the case of Austria).

[3] As of January 1st 2013, the new rules of the FSA, the UK's financial regulator, do not allow financial advisers to receive commission offered by product providers, even if they intend to rebate these payments to the consumer.

[4] For instance, the Dutch authorities have limited initial commission for insurance companies' life and protection business to 50% of total payment. In Denmark and Finland initial commissions on life and pension sales have been banned, e.g., on pension products as early as in 2005 in Finland and in 2006 in Denmark. For some details see the FSA's review of retail distribution conducted in 2007.

[5] Cf. Bebchuck and Fried (2010).

is that the implications must be analyzed under the ensuing optimally (re-)structured compensation.

In our model, we identify three effects that a mandatory deferral has on a bank's (principal's) incentives to induce its agents to exert higher or lower diligence. We review these effects next and then show how, taken together, they provide guidance on when such regulation can have the intended positive effect and when it risks backfiring.

The first effect is indeed positive and arises from the fact that in our model it may be optimal without regulation to pay a bonus that is not contingent on future performance of the business and thus only incentivizes the task of acquisition and growth. When regulation forces the bank to delay any bonus, it becomes optimal for the bank to make incentives that previously were targeted exclusively at the acquisition task now also contingent on performance. While this restructuring of incentives increases the *level* of compensation cost for the bank, it will lower the *marginal* cost of inducing diligence. Thus, when performance insensitive upfront payments are made, mandatory deferral unambiguously increases equilibrium diligence. A second positive effect arises when regulation is particularly restrictive. We find that in this case regulation may virtually force the bank to induce higher diligence as a lower level of induced diligence and, thereby, a lower agency rent would not be compatible with the additional objective of inducing acquisition effort. Still, there is also a negative effect of regulation, which arises when the acquisition constraint is slack. In this case, the bank would never find it privately optimal to make use of performance-insensitive pay. When the regulator now imposes a further delay of compensation, regulation will not just increase the *level* of compensation cost, but also increase the *marginal* cost of inducing diligence. In equilibrium, mandatory deferral of the bonus may then lead, via this effect, to lower rather than higher diligence.

Our analysis thus highlights that the effects of mandatory deferral are more subtle than a simple, but misleading argument along the lines of "regulation increases the price of diligence, so people "consume" less diligence." This simplistic argument neglects the important difference between overall compensation cost and marginal compensation cost. While overall compensation cost must increase when regulation binds, the effect on marginal cost is a priori unclear. In fact, our results clearly show that the "price of diligence" might actually decrease as a result of mandatory deferral. Due to this theoretically ambiguous effect of regulation, it is necessary to solve a concrete model analyzing a concrete policy to understand the role of economic primitives in shaping marginal compensation cost.

One robust insight of our model is the following. We conduct a comparative analysis of the impact of mandating a longer deferral of contingent compensation in terms of the incentives for acquisition (of deals or customers) that are given by the bank to its agents. When these acquisition incentives are high, a binding mandatory deferral of incentive pay will lead to higher equilibrium diligence. Intuitively, if the agent has already significant money at stake (unrelated to the compensation for diligence effort) the threat of losing it all (if the bad outcome is observed early) makes it cheaper to provide incentives for higher diligence effort. In contrast, when acquisition incentives are low, we find that it is optimal not to impose any such restrictions. Also, mandatory deferral risks backfiring when the bank has already high incentives to induce diligence, e.g., as the risks are largely borne by the bank itself rather than, for instance, the holders of securities through which it has off-loaded these risks. This also applies when the task of acquisition is separated from that of exerting diligence in screening and managing risks, while mandatory deferral should lead to higher equilibrium diligence when the provision of acquisition incentives becomes more onerous for the bank as, for instance, competition intensifies. Finally, our results point to interesting interaction effects between other regulatory policies targeted at improving bank's incentives and mandatory deferral. In particular, if higher capital requirements increase the bank's own incentives for diligence, imposing additional mandatory deferral constraints is more likely to backfire.

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.[6] In Thanassoulis (2012) competition for bankers drives up market levels of remuneration and, thereby, increases banks' default risk. In Acharya and Volpin (2010), high pay is a sign of weak governance, which drives up compensation costs at other banks and may induce also their shareholders to implement a weak governance system (cf. also Dicks, 2012). In Bénabou and Tirole (2013) competition for agents not only

[6] According to the managerial entrenchment view advanced by Bebchuk and Fried (2004) executive compensation is set by CEOs themselves rather than boards on behalf of shareholders (see also Kuhnen and Zwiebel, (2009)). Here, regulation would be beneficial to shareholders. Empirically, there seems to be little evidence that those banks where interests of top management were better aligned with shareholders' interests performed better. (For some evidence to the contrary, see, for instance, Fahlenbrach and Stulz (2011)).

changes the size of compensation, but it can inefficiently affect the structure of compensation, inducing an excessive reliance on high-powered performance pay.[7] Inderst and Pfeil (2013) consider compensation regulation when there is a tension between the task of generating loan prospects and that of screening out bad loans. Instead, in our model the considered multiple tasks are, in fact, complementary: When agents are induced to exert more diligence, the resulting higher rent also creates positive incentives to generate more business opportunities.[8] Also, we allow payments at any point of time, thereby endogenizing the timing of bonus payments, rather than restricting compensation to only immediate payments or at most one additional period, respectively. The optimal mix of short- and long-term compensation for corporate executives is analyzed in Peng and Roell (2011), Chaigneau (2012), and Edmans et al. (2012). There, early vesting allows to reduce compensation risk for risk-averse managers or to smooth their consumption over time.[9] In our model, the manager is risk-neutral but (relatively) impatient, while deferral increases the informativeness of the performance measure.[10]

The key task of diligence in our model is aimed at reducing the likelihood with which a possibly rare but observable (and for the bank, its customers, or society critical) event will occur. Our modeling of such a negative event is shared with Biais et al. (2010) and notably Hartman-Glaser et al. (2012) as well as Malamud et al. (2013). In fact, we can rely on the technical analysis of the latter papers and restrict our optimal contract design problem to the determination of bonus payments made at most at countably many points of time. The focus in this paper is, however, on the implications that regulation has on optimal incentive compensation and, thereby, on the equilibrium provision of diligence. Section 2 introduces the baseline model. At the end of this section we also lay out the roadmap for the further analysis.

[7]Other papers, such as Bolton et al. (2010), have advanced the idea to incorporate features of debt into bank managers' compensation so as to reduce risk-taking incentives.

[8]More specifically, while our model interacts two costly tasks, Inderst and Pfeil (2013) combine an *ex-ante* moral-hazard problem with a problem of *interim* private information, in the spirit of the larger literature on "delegated expertise" (c.f. e.g., Levitt and Snyder (1997) or, more recently, Gromb and Martimort (2007), as well as the application in Heider and Inderst (2012)).

[9]Early vesting may also enhance project choice, as pointed out in Briseley (2006) and Laux (2012).

[10]In the model of Chaigneau (2012) the stock price becomes noisier over time, which limits its informativeness (though the information from the whole history of stock prices is still increasing over time).

2 Model Setup

To evaluate the positive implications of mandatory regulatory deferral of bonus payments, we set up a broadly applicable principal-agent model in which the timing of pay is central to the relationship. The agent has two tasks: that of generating business opportunities, to which for simplicity we refer to as (deal or customer) *acquisition*, and that of exerting sufficient *diligence* in selecting or managing these opportunities. As in most of the literature, the agent performs these tasks only once at the very beginning of the model.[11] The key distinction between the two tasks is that acquisition is immediately observable by the principal, whereas information about diligence is continuously revealed over time (giving rise to endogenous timing of pay). We want to think about the principal as being a financial institution, whereas the agent in our model can be interpreted in various ways: a senior executive, a loan officer, or a broker of financial products (also see the particular application of loan defaults with externalities in Section 5).

Through exerting unobservable effort a at private disutility $k(a)$, the agent generates an opportunity with probability a. Subsequently, through exerting unobservable diligence μ at private disutility $c(\mu)$ the agent can affect the likelihood with which a - relatively infrequent - "bad event" occurs or is avoided.[12] Only over time, the principal can learn about diligence through the absence of such an event. Formally, we let μ represent the probability with which the occurrence of this event is exponentially distributed with parameter λ_L instead of with parameter $\lambda_H > \lambda_L$. Whether such an event occurred, as well as whether an opportunity was acquired in the first place, are both verifiable events and thus contractible. This setting thus encompasses scenarios in which lack of diligence might only be exposed with considerable delay or only in extreme times, a feature that seems to be relevant in many applications in the financial sector, in particular when diligence affects risk-taking.

Diligence μ represents a continuous variable, which will ensure that it reacts also to marginal changes in compensation and, thereby, regulation. The respective cost function

[11] With respect to the literature on compensation regulation, see for instance recently Peng and Roell (2011) or Benabou and Tirole (2013). Instead, Edmans et al. (2012) consider repeated effort choice in a model of CEO compensation.

[12] Note that the two tasks are thus not conflicting as a high value of a or μ does not change (or, in particular, increase) the marginal cost of the other task. Such an interdependence could however be easily incorporated into our analysis, provided it is not too strong. If high effort in one task raises the cost of the other task by "too much," it would be trivially optimal to split up the tasks among two agents.

$c(\mu)$ is twice continuously differentiable. To obtain an interior solution, we stipulate that $c''(\mu) > 0$, $c'(\mu) = 0$ for $\mu = 0$, and that $c'(\mu)$ becomes sufficiently large as $\mu \to 1$. For simplicity, we consider two levels of acquisition effort, $a_h > a_l > 0$, with respective disutility $k_h > k_l$.[13]

As for the principal's revenue, we only require that $a\pi(\mu)$, the expected present value of profits gross of compensation costs, is continuously differentiable and increasing in μ. We will provide an example showing how $\pi(\mu)$ can be derived from model primitives in Section 5.1. There we will also introduce a negative externality upon the occurrence of the "bad event", which the principal and its agent do not fully internalize. For now, it is only relevant to stipulate that the expected present value of this externality, $\omega(\mu) > 0$, is continuously differentiable and decreasing in μ, creating scope for regulation in the first place.

Both parties are risk neutral and discount payoffs at respective discount rates $r_A > r_P$, implying that the agent is relatively more impatient than the bank (the principal). The assumption of relative impatience makes it costly for the principal to delay bonus payments. This assumption is common in the literatures on labor, executive compensation, and contracting (cf. Rogerson 1997, DeMarzo and Duffie 1999, or DeMarzo and Sannikov 2006).[14] It is justified on various grounds. For instance, employees may have higher liquidity preferences than the bank does, as they are (more) credit-constrained. Note that the agent is also unable to borrow against his future (expected) income, as this would undermine his incentives and, thereby, his future ability to repay such a loan.

Compensation Contracts. Compensation payments must be non-negative and can be conditioned on all information available at the time of payment, i.e., more formally, they are adapted to the filtration generated by Y_t where $Y_t = 1$ indicates that the bad event has occurred before time t, and $Y_t = 0$ that it has not. Note that while this prima facie precludes claw-back clauses, this is not the case as long as the respective payments can not yet be consumed by the agent until these clauses expire. We allow payments to be made at any point in time. As we consider an open time horizon, say rather than two periods only, we will be able to pin down (long-term) bonus payout times through a first-

[13] This restriction is mainly for ease of exposition. See Appendix B for an analysis of the model with continuous acquisition effort.

[14] An alternative (but less tractable) way of making incentive compensation costly would be to assume risk-aversion on the side of the agent.

order condition. Further, we can analyze how outcomes change when regulation imposes a (marginally) longer deferral period, as motivated by existing regulatory proposals.[15] In particular, we consider the impact of a regulation that requires all compensation to be made contingent on all performance-relevant information available at the time of pay and that such contingent compensation can not be paid out before a certain time $\tau > 0$. From the principal's perspective, this regulatory intervention is simply an exogenous restriction to the contract design space, trivially increasing the *level* of compensation cost.[16]

Organization. The analysis of this model now proceeds in three logical steps. First, in Section 3 we study the unregulated benchmark economy (or equivalently the special case $\tau = 0$). This allows us to understand the important trade-offs relevant for optimal contract design (see Section 3.1) and the role of primitives for the unregulated equilibrium outcome (Section 3.2) in the absence of the confounding effects of regulation. In the subsequent and main Section 4, we use the identical structure to understand the effects of regulation on compensation design (Section 4.1) and the equilibrium provision of diligence (Section 4.2). Since this positive analysis is motivated by existing regulation this section still treats the deferral time τ as exogenous. Finally, in Section 5 we close our model for a particular application (loan defaults with externalities) that implies a specific regulatory objective function generating the scope for regulation, i.e., the endogenous choice of τ. All proofs are collected in Appendix A. Note that some of the proofs for the special case ($\tau = 0$) are contained in the proofs of the more general Propositions in Section 4. Finally, Appendix B contains some additional material for the case where also acquisition effort is a continuous variable.

3 Equilibrium without Regulation

3.1 Optimal compensation contracts

Following the technical results of Malamud et al. (2013), it is without loss of generality in our risk-neutral setting to restrict attention to payment schemes that specify a countable number of times T_i at which payments b_i are made. Optimally, these will be made if and

[15] The explicit objective of current regulatory proposals along these lines is to induce the exertion of greater diligence by the respective agents, e.g., to avoid excessive risk-taking or unsuitable advice.

[16] Of course, under optimal regulation the value of τ is chosen endogenously. We defer this discussion until Section 5 where we derive the objective function of the regulator for a particular application.

only if acquisition was successful and the "bad event" has not occurred by the respective time T_i, i.e., if $Y_{T_i} = 0$.[17] Note however, that in the absence of regulation a positive payment b_0 in $T_0 = 0$ may occur before the bad event can even occur with positive probability.

Take now a given choice of diligence μ and acquisition effort a. The agent's discounted expected payoff equals

$$V_A(a,\mu) = a\left(\sum\nolimits_i b_i e^{-r_A T_i}\left[\mu e^{-\lambda_L T_i} + (1-\mu)e^{-\lambda_H T_i}\right] - c(\mu)\right) - k(a). \tag{1}$$

Recall that the agent discounts compensation with the rate r_A. Also, the costs of diligence, $c(\mu)$, are only incurred when acquisition was successful. Through affecting whether a "bad event" occurs with arrival rate λ_L rather than λ_H, higher diligence makes it more likely that the agent will receive any positive compensation $b_i > 0$ that is delayed when $T_i > 0$.

To ensure that μ is indeed optimal for the agent, the corresponding first-order condition must be satisfied:

$$\sum\nolimits_i b_i e^{-r_A T_i}\left(e^{-\lambda_L T_i} - e^{-\lambda_H T_i}\right) = c'(\mu). \tag{2}$$

Given that the left-hand side is non-negative, there is indeed a unique μ that solves the first-order condition, which is also sufficient as $c'' > 0$. To induce high acquisition effort, $a_h = 1$, it must hold, in addition, that $V_A(a_h,\mu) \geq V_A(a_l,\mu)$:

$$\sum\nolimits_i b_i e^{-r_A T_i}\left[\mu e^{-\lambda_L T_i} + (1-\mu)e^{-\lambda_H T_i}\right] - c(\mu) \geq \frac{k_h - k_l}{a_h - a_l} =: k. \tag{3}$$

In what follows, it is convenient to stipulate that $a_h = 1$, which abbreviates some expressions, though the subsequent comparative analysis will depend only on k. The total cost of compensation to the bank then equals

$$W = \sum\nolimits_i b_i e^{-r_P T_i}\left[\mu e^{-\lambda_L T_i} + (1-\mu)e^{-\lambda_H T_i}\right], \tag{4}$$

which now uses the bank's (the principal's) discount rate r_P. We define $\Delta_r = r_A - r_P > 0$, which captures the loss from delaying compensation as the agent is more impatient than the principal. Also, denote $\Delta_\lambda = \lambda_H - \lambda_L$. As this is the difference in the respective

[17] The restriction is thus to rule out rates at which payments can be made continuously. In Appendix C of the working paper we appeal to results in Malamud et al. (2013) and argue that such rates are indeed not optimal in our setting with risk neutrality.

rates with which a bad event occurs, it captures the speed of learning.[18] The tension between a loss from delay due to differences in impatience, as captured by Δ_r, and higher incentives through the use of more information, as captured by Δ_λ, represents the key trade-off in the compensation design problem.

For a given level of diligence, the bank's program is to minimize compensation costs W subject to the incentive constraints (2) and (3) as well as the non-negativity constraints $T_i \geq 0$ and $b_i \geq 0$. For the subsequent characterization the following observations are now helpful. Choosing $b_0 > 0$ relaxes only the acquisition constraint (3), but not the diligence incentive constraint (2). Further, suppose for a moment that there is only a single delayed bonus $b_i > 0$ paid at $T_i > 0$, which will indeed hold in equilibrium. With a slight abuse of notation, call this the long-term bonus $b_i = b_T$ paid at $T_i = T$. Then, regardless of the choice of b_0, from (2) this must satisfy

$$b_T = c'(\mu)\,\frac{e^{(r_A+\lambda_H)T}}{e^{\Delta_\lambda T}-1}. \tag{5}$$

We obtain the following characterization:[19]

Proposition 1 *To implement a given level of diligence μ, together with high acquisition effort, at lowest cost of compensation, the bank chooses a single, uniquely determined long-term bonus b_T, which satisfies (5), and a unique timing T. If*

$$\mu < \frac{1}{2}\left(1-\frac{\Delta_r}{\Delta_\lambda}\right) \tag{6}$$

holds and the costs of acquisition effort satisfy $k > \overline{k} > 0$, an additional up-front bonus $b_0 > 0$ is paid:

$$b_0 = k + c(\mu) - c'(\mu)\left(\mu + \frac{1}{e^{\Delta_\lambda T}-1}\right). \tag{7}$$

Proof. See Appendix A.

Before we comment on this characterization, note that it applies equally when, instead of stipulating a binary acquisition effort, a was continuous. Then, in equation (7), k would be replaced by the respective marginal cost, $k'(a)$, evaluated at the acquisition level that

[18] Of course, as is standard in problems of moral hazard, along the equilibrium path there will not be any learning about the chosen strategy.

[19] Hartman-Glaser et al. (2012) analyze a similar model with binary diligence effort. However, they impose a parameter restriction that essentially implies that (6) does not hold, so that an up-front payment is never optimal (cf. also the discussion in Appendix C of our working paper version).

the bank wishes to implement (cf. Appendix B for a discussion). Take now condition (6), which is crucial for whether, in addition to the long-term bonus b_T, there will also be an up-front bonus b_0, in which case the bank's objective actually becomes that of maximizing joint surplus. For given μ, this reduces to the problem of minimizing the deadweight loss that arises from deferring compensation, given that the agent is more impatient than the bank. When high acquisition effort is chosen, the joint surplus of the bank and the agent is $\pi(\mu) - c(\mu) - k - D$, where

$$D = b_T \left(e^{-r_P T} - e^{-r_A T}\right) \left[\mu e^{-\lambda_L T} + (1-\mu)e^{-\lambda_H T}\right]$$

is the deadweight loss from delay due to relative impatience. After substituting from (5), we have

$$D = c'(\mu)\left(e^{\Delta_r T} - 1\right)\left[\mu + \frac{1}{e^{\Delta_\lambda T} - 1}\right], \tag{8}$$

which is indeed zero when $\Delta_r = 0$ as both parties discount the future at the same rate, $r_A = r_P$.

For this problem, i.e., to minimize deadweight loss, we need to distinguish between two cases. In the first case, there is a unique interior value of delay T at which deadweight loss is minimized. In particular, reducing delay further would then require to push up the bonus too much so as to still preserve incentives, and the overall deadweight loss would increase rather than decrease. This case applies precisely when condition (6) holds: That is, when i) Δ_r/Δ_λ is relatively low, i.e., when the costs from delaying the bonus, as captured by the difference in the respective discount rates Δ_r, are small compared to the gain in information, as captured by the difference in the arrival rates Δ_λ; and when ii) the level of diligence that the bank wants to implement is relatively low, as high diligence would require, ceteris paribus, a high long-term bonus and would thus make delay more costly. This case is further illustrated in Figure 1. There, for the presently discussed case the solid line depicts deadweight loss as a function of the chosen timing of the long-term bonus, where the latter must be adjusted so as to preserve the agent's incentive to choose a given μ.

When condition (6) does not hold, deadweight loss from delay would always become strictly lower as T decreases.[20] This case is depicted by the dotted line in Figure 1. To preserve incentives, however, this would require to pay an always higher long-term bonus

[20] In this case, the minimium of deadweight loss would thus be obtained for $T \to 0$ and would be equal to $c'(\mu)\,\Delta_r/\Delta_\lambda$.

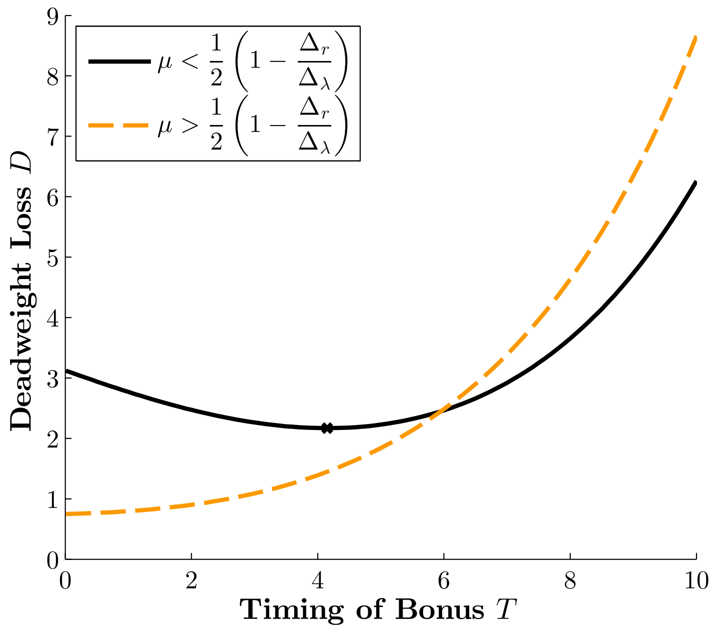

Figure 1. This graph plots "deadweight loss" as a function of the timing of the long-term bonus T. The solid line plots the case in which an interior minimum obtains, i.e., here for the specification $\mu = 0.05 < \frac{1}{2}\left(1 - \frac{\Delta_r}{\Delta_\lambda}\right)$ and $c(\mu) = 25\mu^2$. The dotted line plots the case when $\mu = 0.3 > \frac{1}{2}\left(1 - \frac{\Delta_r}{\Delta_\lambda}\right)$ and $c(\mu) = \mu^2$. For both cases, we specify $\Delta_r = 0.3$, $\Delta_\lambda = 0.6$.

as T decreases. Then, the agent's acquisition constraint will become slack, so that we are no longer in the presently analyzed regime (where the objective of the bank coincides with joint surplus maximization).

At this point it is useful to note that, making use of the first-order condition for diligence (5), an agent's ex-ante payoff gross of acquisition effort costs k can be decomposed as follows. Suppose for a moment that the agent faces only the task to exert diligence effort and that it is immediately *observed* whether λ_L or λ_H was realized. It is then optimal to pay an immediate bonus $b = c'(\mu)$ upon observing λ_L (which occurs with probability μ), so that the agent's rent equals $c'(\mu)\mu - c(\mu)$. Returning now to our original problem, where this is not observed, this is also the limit when the agent's long-term bonus is always further delayed, while otherwise the agent's payoff increases by $c'(\mu)/[e^{\Delta_\lambda T} - 1]$. The difference between the thereby increased "diligence rent" and the costs of acquisition effort k yields b_0 in expression (7), which is the up-front bonus that is paid additionally to induce high acquisition effort.

Additional Characterization. We continue by providing additional details for the characterization obtained in Proposition 1. These follow as well from the proof of Proposition 1 in the Appendix. In particular, Lemma 1 introduces three different regimes, which are obtained as k gradually increases. They differ intuitively depending on whether the acquisition constraint binds and, in addition, also on whether from condition (6) an up-front bonus will be optimal or not. When we subsequently refer to these regimes in the main text, however, we will not need to make use of the partially explicit characterization of the optimal compensation that is obtained in Lemma 1. Still, we state it for completeness and also to make more transparent the subsequent comparative statics analysis.

Lemma 1 *The delay T of the long-term bonus in the characterization of the optimal compensation (Proposition 1) is obtained as follows in three different regimes: For $k < \underline{k}$ the acquisition constraint (3) is slack (Regime 1), there is no up-front bonus ($b_0 = 0$) and T is given by*

$$T_1 = \frac{1}{\Delta_\lambda} \ln \left(1 + \frac{\Delta_\lambda - \Delta_r + \sqrt{(\Delta_\lambda - \Delta_r)^2 + 4\Delta_\lambda \Delta_r \mu}}{2\Delta_r \mu} \right). \tag{9}$$

For $k \geq \underline{k}$, where the acquisition constraint binds, there are two cases to distinguish. In Regime 2 there is no up-front bonus ($b_0 = 0$) and T is given by

$$T_2 = \frac{1}{\Delta_\lambda} \ln \left(1 + \frac{c'(\mu)}{k + c(\mu) - c'(\mu)\mu} \right). \tag{10}$$

This applies when either condition (6) does not hold or always when k is still sufficiently low with $k \leq \overline{k}$. When, instead, (6) holds and $k > \overline{k}$, Regime 3 applies with $b_0 > 0$ and T_3 as the unique positive solution for T to

$$\frac{1 - e^{-\Delta_r T}}{1 - e^{-\Delta_\lambda T}} \frac{1}{1 + \mu \left(e^{\Delta_\lambda T} - 1\right)} = \frac{\Delta_r}{\Delta_\lambda}. \tag{11}$$

The thresholds on acquisition costs satisfy:

$$\underline{k} = c'(\mu) \left(\mu + \frac{1}{e^{\Delta_\lambda T_1} - 1} \right) - c(\mu), \tag{12}$$

$$\overline{k} = c'(\mu) \left(\mu + \frac{1}{e^{\Delta_\lambda T_3} - 1} \right) - c(\mu), \text{ for } \mu < \frac{1}{2} \left(1 - \frac{\Delta_r}{\Delta_\lambda} \right). \tag{13}$$

Further Discussion and Comparative Analysis. The further characterization of the delay of the long-term bonus in Lemma 1 gives now rise to an immediate comparative result on the duration of optimal compensation, which includes both the size and the timing of all payments. As can be seen immediately from the respective expressions, the timing of the long-term bonus T is independent of acquisition costs k in regimes 1 and 3, i.e., when $T = T_1$ or $T = T_3$, while $T = T_2$ strictly decreases with k in regime 2. Note, that the optimal compensation plan in regime 3 implies an additional upfront bonus b_0. We have from (7) that the up-front bonus b_0 increases one-for-one with k, while b_T remains unchanged.

The characterization of regimes in Proposition 1 and Lemma 1 is in terms of acquisition cost k treating the choice of diligence μ as exogenous. We now provide the corresponding characterization in terms of μ. For this note first that the threshold levels $\underline{k}$ and $\overline{k}$ for regimes 1 and 3, as stated in Lemma 1, are strictly increasing functions of μ in the relevant parameter region, so that the inverse functions are well defined. We can thus obtain the following immediate Corollary.

Corollary 1 *Consider a given level of acquisition cost $k > 0$. Then, one of the following regimes from Proposition 1 and Lemma 1 applies, depending on the bank's choice of implemented level of diligence μ: Regime 3 applies when μ is low (provided that $\Delta_r < \Delta_\lambda$), regime 2 applies for intermediate levels, and regime 1 applies for high levels.*

Taken together, we are thus most likely to be in regime 1 when either the acquisition task requires little effort costs or when the bank wants the agent to exert high diligence effort (and the diligence rent is therefore high). On the other hand, provided that $\Delta_r < \Delta_\lambda$, there will be an up-front bonus next to a long-term bonus (regime 3) when the bank wants to induce relatively little diligence effort but when the acquisition task is sufficiently important as k is high. We illustrate these insights in Figure 2 for an example which features $\Delta_r < \Delta_\lambda$.

Finally, we can determine the comparative statics of the optimal bonus times in μ using the different regimes described in Corollary 1.

Corollary 2 *An increase in the level of diligence μ leads to a strict reduction in the delay of the long-term bonus in regimes 1 and 3 of Proposition 1 and Lemma 1, but to a strict increase in delay in regime 2. Also, in regime 3, where an up-front bonus $b_0 > 0$ is paid, this bonus strictly decreases.*

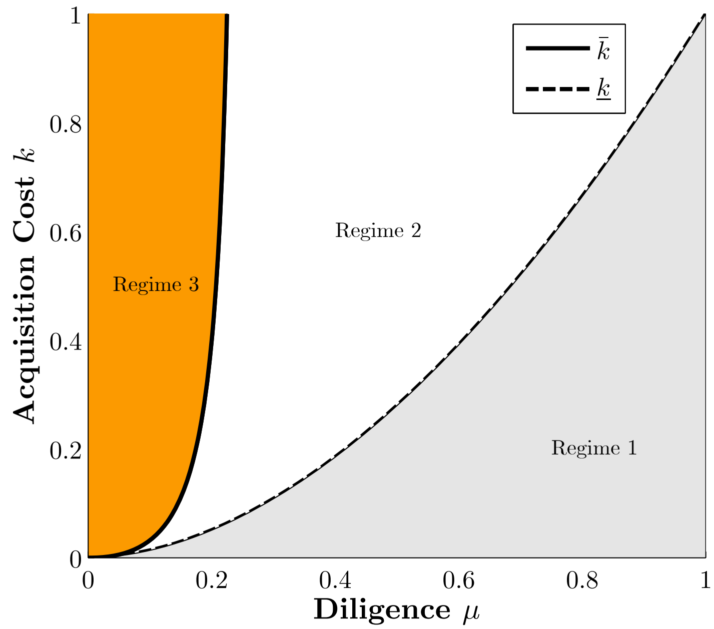

Figure 2. This graph plots the (μ, k) combinations that give rise to the three regimes described in Proposition 1. The cost function for diligence satisfies $c(\mu) = 0.5\mu^2$. The remaining parameters are $\Delta_r = 0.3$, $\Delta_\lambda = 0.6$.

Proof. See Appendix A.

Corollary 2 suggests that overall we may not observe a monotonic relationship between diligence and thus the frequency of "bad events," such as loan defaults or customer complaints, and the importance and timing of deferred pay for the responsible agents. For an illustration of this non-monotonicity, Figure 3 depicts the equilibrium choice of delay T as a function of the implemented level of diligence μ.

For our further discussion the comparative results in regimes 1 and 3, in particular, will be important. While in regime 2 the timing and size of the long-term bonus are essentially tied down rather mechanically by the two binding constraints, we already noted that in regimes 1 and 3 they are determined from the respective first-order conditions, namely that of maximizing only the bank's profits (in regime 1) and that of maximizing the joint payoff (in regime 3). As the agent's expected deferred compensation must increase when a higher diligence shall be induced, which would ceteris paribus increase deadweight loss given the difference in the bank's and the agent's discount rate, it becomes optimal to pay the long-term bonus earlier.

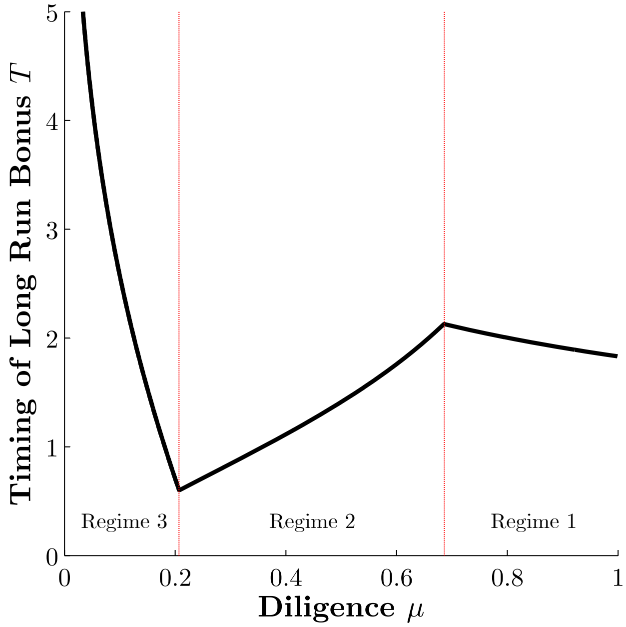

Figure 3. This graph plots the equilibrium choice of delay T as a function of the implemented level of diligence μ. As in Figure 2 the cost function for diligence satisfies $c(\mu) = 0.5\mu^2$. The remaining parameters are $\Delta_r = 0.3$ and $\Delta_\lambda = 0.6$. The cutoffs for the respective regimes follow directly from Figure 2 for $k = 0.5$.

3.2 Equilibrium Provision of Diligence

From the characterization of the optimal contract in Proposition 1 and Lemma 1 we can obtain for any given μ the minimum compensation costs. We denote these by $W(\mu)$ and defer a full characterization to the proof of Proposition 2. The equilibrium level of diligence effort is then obtained from maximizing bank profits net of compensation costs

$$\Pi(\mu) = \pi(\mu) - W(\mu).$$

One can show that $W(\mu)$ is everywhere continuously differentiable. As long as we can abstract from corner solutions, an optimally implemented μ^* thus solves the first-order condition

$$\pi'(\mu^*) = W'(\mu^*).$$

Without additional restrictions on functional forms, μ^* may not be pinned down uniquely.[21] For convenience only, we suppose in what follows, however, that $\Pi(\mu)$ is strictly quasi-

[21] One such restriction is that, next to $\pi'(\mu) > 0$ and $\pi''(\mu) \leq 0$, the marginal costs of effort $c'(\mu)$ are sufficiently convex, i.e., that $c'''(\mu)$ is everywhere sufficiently high.

concave.[22] We denote the respective contractual parameters that arise from Proposition 1 for $\mu = \mu^*$ by T^*, b_T^*, and b_0^*. To conclude the characterization, we show that, as is intuitive from the previous observations, even when accounting for the equilibrium choice of μ, the characterization of regimes from Proposition 1 and Lemma 1 in terms of k thresholds survives. The only difference is that now the thresholds for k must be defined while using the respective equilibrium choice of μ (see proof of Proposition 2). In particular, as with a higher k it becomes more expensive to incentivize the acquisition task, the equilibrium moves from regime 1 to regime 3, provided that the now modified condition (6) holds so that indeed $b_0^* > 0$ for high k.

Proposition 2 *At the bank's optimal choice of diligence μ^*, we have the following characterization result, making use of the three regimes introduced in Proposition 1 and Lemma 1. For $k < \underline{k}^*$ regime 1 applies; for $k > \overline{k}^*$ and $\Delta_r/\Delta_\lambda < \gamma$ with $\gamma > 0$, regime 3 applies; otherwise, regime 2 applies.*

Proof. See Appendix A.

4 Equilibrium with Deferral Regulation

In this section, we consider the impact of a mandatory deferral of incentive pay on compensation design and the level of diligence that the bank optimally induces. This analysis is motivated by the following two observations (cf. the Introduction). Such a policy is frequently considered and applied at various levels in the financial industry, ranging from executive compensation to the structure of commissions at the front end. Moreover, the explicit objective of such a policy is to induce the exertion of greater diligence by the respective agents, e.g., to avoid excessive risk-taking or unsuitable advice. We wish to analyze when, in our model, this holds true.

Recall that the considered regulation requires that all compensation must be made contingent on subsequent performance and can not be paid out before a certain time $\tau > 0$. Since compensation regulation targets the agent's principal, the question is how mandatory deferral induces the bank to restructure incentives. Of course, any such restrictions on compensation design must by construction weakly increase the *level* of

[22] In the working paper version, we did not make this assumption, though. All subsequent results are there shown to hold also without this assumption (namely, by then stating our assertions in terms of comparative statics over the respective sets).

compensation cost to implement a given μ, since the bank already minimizes compensation cost in the absence of such restrictions. But this is, by itself, inconsequential for the the bank's incentives to induce a higher or lower level of diligence, as these depend on the *marginal* compensation cost. In what follows, we derive conditions for when the marginal cost of inducing (higher) diligence is higher or lower under mandatory deferral, so that in equilibrium regulation decreases or increases diligence.

Our analysis proceeds as follows. First, we characterize in section 4.1 the optimal compensation choice under regulation. Subsequently, in section 4.2 we ask what level of diligence the bank wants to optimally induce under regulation. Our procedure thus mirrors the steps of the analysis without regulation. Subsequently, we make use of these results to analyze a regulator's optimal choice of deferral (Section 5).

4.1 (Constrained) Optimal Compensation Contracts

The characterization of the optimal compensation contract under regulation follows essentially the same principles as that without regulation in Proposition 1. Still, depending on the size of the minimum deferral time τ, we now have to make additional case distinctions, which slightly complicates the exposition. As a consequence, we proceed stepwise.

The first thing to note is that now a given level of diligence may no longer be implementable at all. This follows from the following reasoning. Recall that $\mu c'(\mu) - c(\mu)$ would be the agent's payoff in a suitably adjusted stationary model where it was known immediately whether λ_L or λ_H was realized. In our model, this is also the agent's payoff in the limit when the diligence level μ is implemented through a single bonus that is always longer delayed, $T \to \infty$, while $b_0 = 0$. When this falls short of k, which is the expected payoff required to incentivize acquisition effort, the respective level of μ (and, intuitively, all lower levels; cf. also below) cannot be implemented when the required delay of any bonus, τ, is sufficiently large. Put differently, then the agent's diligence rent alone is not sufficiently high to incentivize acquisition. This result already points to a rather immediate effect of regulation, which we further discuss in the next section.

Lemma 2 *Suppose $k > \mu c'(\mu) - c(\mu)$ for some level of diligence μ. Then, under mandatory deferral of a bonus until at least time τ, the respective diligence level μ can only be implemented when $\tau \leq T_2$, with T_2 given by (10), which is strictly decreasing in k and strictly increasing in μ. Likewise, for given τ there is a unique threshold $\underline{\mu}(\tau) > 0$, such* *that only diligence levels $\mu \geq \underline{\mu}(\tau)$ are feasible. The minimum diligence level $\underline{\mu}(\tau)$ that this requires is a strictly increasing function of τ and k. When $k \leq \mu c'(\mu) - c(\mu)$ any μ can be implemented, regardless of τ.*

Proof. See Appendix A.

For future reference it is helpful to make the following additional definition. Define for the case where the minimum deferral time goes to infinity $\overline{\mu} = \lim_{\tau \to \infty} \underline{\mu}(\tau)$, which is implicitly characterized by:[23]

$$k + c(\overline{\mu}) - c'(\overline{\mu})\overline{\mu} = 0. \tag{14}$$

Implicitly differentiating (14) reveals that $\overline{\mu}$ is strictly increasing in the respective cost of acquisition effort k. (Note that this follows also from the observation in Lemma 2 that $\underline{\mu}$ is strictly increasing in τ and k: As the expected compensation that the agent must obtain from acquisition increases, the respective bonus that is paid at a particular point in time τ must increase as well, which induces higher diligence.) The following characterization of the optimal contract to implement a given level of diligence μ under regulation, provided that this is feasible as $\mu \geq \underline{\mu}(\tau)$ (cf. Lemma 2), follows intuitively the same principles as Proposition 1. Note also that we provide further details in a subsequent Lemma (as previously in Lemma 1).

Proposition 3 *Under regulation the cost-minimizing compensation contract consists of at most two payments, which occur at τ and/or at a uniquely determined $T > \tau$. In particular, if*

$$\frac{\Delta_r}{\Delta_\lambda} < \frac{1 - \mu\left(1 + e^{\Delta_\lambda \tau}\right)}{1 + \mu\left(e^{\Delta_\lambda \tau} - 1\right)} \tag{15}$$

holds and the costs of acquisition effort satisfy $k > \overline{k}(\tau) > 0$, there are two payments b_τ and b_T determined from the binding constraints (2) and (3).

Proof. See Appendix A.

Recall that an up-front bonus $b_0 > 0$ is no longer feasible under regulation. When the mandatory deferral time τ is not too long, it can now be optimal for the bank to make

[23] When τ becomes too high, however, the bank's profits from this line of business will become negative, which - as discussed below - should impose a (participation) constraint on regulation.

two long-term bonus payments: one at τ and one strictly later at some time $T > \tau$. We can show in this case that when regulation becomes gradually more severe as τ increases, T shrinks and with it the distance between the two points of time when the respective bonus payments are made, $T - \tau$. When the required deferral τ becomes sufficiently large, however, then there will always be a single bonus that is paid exactly at the first instance when it is allowed to do so (at τ).

For given induced diligence μ, with regulation condition (15) determines whether, provided that k is not too low, there will be two bonus payments: a short-term payment after the minimum deferral period imposed by regulation and a long-term payment that is further delayed. While not at first evident, the qualitative properties of this condition are analogous to those of the respective condition (6) without regulation: Two bonus payments still are more likely when the induced level of diligence is relatively low and also when Δ_r/Δ_λ is relatively low, i.e., when the costs from delaying incentive pay are small compared to the gain in information.

Additional Characterization. To bring out more transparently the analogy to the characterization without regulation, as in Lemma 1 we next provide some additional details for the characterization, which follow as well from the proof of Proposition 3 in the Appendix.

Lemma 3 *Under regulation, the optimal delay of long-term bonus payments is characterized as follows in three different regimes. If $k < \underline{k}$, with $\underline{k}$ given by (12), the acquisition constraint is slack (Regime 1) and there is a single payment b_T satisfying (5). The optimal delay of the long-term bonus is uniquely determined from $T = \max\{T_1, \tau\}$, where T_1 is given by (9). If $k \geq \underline{k}$ and condition (15) holds, then there exists a threshold*

$$\overline{k}(\tau) = c'(\mu)\left(\mu + \frac{1}{e^{\Delta_\lambda T_3(\tau)} - 1}\right) - c(\mu),$$

such that for $k > \overline{k}(\tau)$ there are two payments b_τ and b_T determined from the binding constraints (2) and (3) (Regime 3). Consequently, the optimal payout times are τ and $T = T_3(\tau) > \tau$, which is the unique solution $T > \tau$ to

$$\frac{1 - e^{-\Delta_r(T-\tau)}}{1 - e^{-\Delta_\lambda(T-\tau)}} \frac{1 + \mu\left(e^{\Delta_\lambda \tau} - 1\right)}{1 + \mu\left(e^{\Delta_\lambda T} - 1\right)} = \frac{\Delta_r}{\Delta_\lambda}. \tag{16}$$

Finally, if either (15) is violated or $k \leq \overline{k}(\tau)$, then there is again a single payment b_T satisfying (5), which now occurs at $T = T_2$, as defined in (10) (Regime 2).

4.2 Equilibrium Diligence under Deferral Regulation

With regulation, the bank's overall problem is the following. The bank still maximizes the respective objective function $\pi(\mu) - W(\mu)$, where the compensation cost function $W(\mu)$ is now obtained from substituting the optimal contract obtained in Proposition 3 and the subsequent Lemma 3. Note that we presently still assume that it is optimal for the bank to induce high acquisition effort. Again, as in the case without regulation, we simplify the analysis by supposing that the now constrained problem has a unique solution. For this we make the dependency on regulation explicit: The bank's optimal choice of induced diligence for a given minimum deferral time τ is denoted by $\mu^*(\tau)$. By allowing for $\tau = 0$ this includes our previous characterization without regulation.[24] It is immediate that regulation operates via its effect on marginal compensation cost: While contractual restrictions must (weakly) increase the overall cost level, our paper stresses the non-trivial effect on marginal cost, which in turn determines the effect on equilibrium diligence.

It is now helpful to consider first separately the two cases where either originally, i.e., without regulation, the acquisition constraint did not bind (regime 1) or where it did bind (regimes 2 and 3). Recall that an up-front bonus $b_0 > 0$ could arise when the acquisition constraint did bind (namely in regime 3).

Case: Slack Acquisition Constraint without Regulation ($k < \underline{k}^*$). When acquisition is relatively unimportant for incentive provision as the respective cost k are sufficiently low, mandatory deferral has a non-monotonic impact on the level of diligence that the bank then optimally wants to implement in equilibrium.

Proposition 4 *Suppose that without regulation regime 1 applies as $0 < k < \underline{k}^*$, so that the acquisition constraint is slack. Regulation has only an effect on the bank's optimal timing of compensation when the minimum deferral time satisfies $\tau > T_1^*$. Then, there*

[24]That is, as nothing is learnt at $\tau = 0$, in this case we allow compensation also to condition on a sale only (which is, $b_0 = 0$).

exists a threshold $\widetilde{\tau} > T_1$ *for the minimum deferral period, so that equilibrium diligence* $\mu^*(\tau)$ *is strictly decreasing in* τ *for* $\tau < \widetilde{\tau}$ *and strictly increasing for* $\tau > \widetilde{\tau}$*, approaching* $\overline{\mu}$ *as* $\tau \to \infty$.

Proof. See Appendix A.

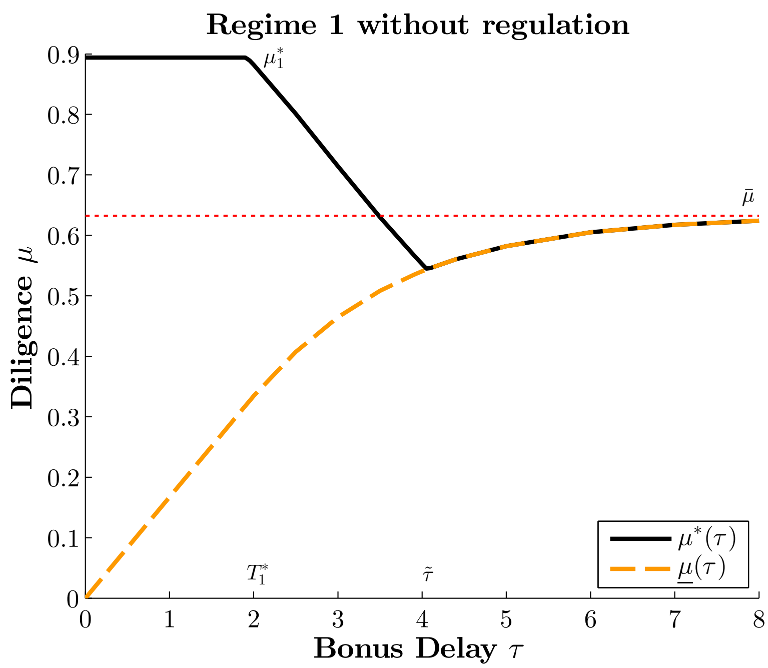

Figure 4. This graph plots the equilibrium diligence μ^* as a function of the mandatory delay of compensation τ for the case when regime 1 obtains without regulation. For small τ, the regulatory constraint does not bind and hence $\mu^* = \mu_1^*$. Once the constraint binds, equilibrium diligence is decreasing in τ until the acquisition constraint of the agent binds at $\tilde{\tau}$, from which point onwards equilibrium diligence is given by the increasing function $\underline{\mu}(\tau)$. The revenue function satisfies $\pi(\mu) = 10\mu$, diligence cost is given by $c(\mu) = \frac{5}{4}c^2$. The remaining parameters are: $k = \frac{1}{2}$, $\Delta_r = 0.3$, $\Delta_\lambda = \lambda_h = 0.6$, and $r_A = 0.2$.

For small mandatory deferral times, i.e., $\tau \leq T_1^*$, regulation does not bind, so that the induced equilibrium diligence choice simple corresponds to the unregulated case $\mu^*(\tau) = \mu_1^*$. Once the regulatory constraint binds and $T_1^* < \tau < \widetilde{\tau}$, the equilibrium level of diligence is decreasing in τ. The intuition for this follows immediately from our earlier comparative result in Corollary 2. There, we observed that without regulation the optimal timing of the long term bonus in regime 1, where the acquisition constraint does not bind, is such that the respective time T_1 strictly decreases with the (then exogenously chosen) level of diligence. Arguing for now only locally, this follows formally from the fact that the cross-derivative of compensation costs with respect to diligence effort and the delay of the long-term bonus is strictly positive. But this implies as well that the incremental costs of inducing a (marginally) higher level of diligence are strictly higher when we (marginally) delay the bonus payment. The proof of Proposition 4 shows that this argument also holds globally, so that for all $\tau > T_1^*$ it holds that, formally, $\frac{d}{d\tau}\left(\frac{dW_1}{d\mu}\right) > 0$. Somewhat surprisingly, our analysis, hence, shows that a mandatory *decrease* in the payout time might actually increase diligence locally.[25] This result clearly shows that our comparative statics results are not driven by generic regulatory interference, but by the concrete policy of mandatory *deferral*.

Since deferral lowers diligence, the agent's diligence rent decreases as the bonus is further delayed until the threshold level $\widetilde{\tau}$ is reached. At $\widetilde{\tau}$ the diligence rent is just high enough to incentivize acquisition. From the threshold $\widetilde{\tau}$ onwards, the acquisition constraint binds and from thereon it is optimal for the bank to set $\mu^*(\tau) = \underline{\mu}(\tau)$, the lowest diligence level that is implementable. As $\underline{\mu}(\tau)$ is strictly increasing in τ, for all $\tau \geq \widetilde{\tau}$ an increase in the mandatory deferral period increases equilibrium diligence up to the limit $\overline{\mu}$. Since in this example the equilibrium level of diligence without regulation, μ_1^*, is greater than $\overline{\mu}$, equilibrium diligence is highest when no (binding) regulation is imposed.[26]

Case: Binding Acquisition Constraint without Regulation ($k \geq \underline{k}^*$). When without regulation regimes 2 or 3 apply, as $k \geq \underline{k}^*$, the impact of regulation is monotonic.

Proposition 5 *Suppose that without regulation regimes 2 or 3 apply as $k \geq \underline{k}^*$. Consider a regulation that constrains the bank's optimal choice of compensation, as $\tau > T_2^*$ in regime 2 or $\tau > 0$ in regime 3. Then, equilibrium diligence under regulation is always strictly increasing as regulation requires a longer delay ($\mu^*(\tau)$ is strictly increasing).*

Proof. See Appendix A.

There are two forces at work that determine the positive impact of regulation in Proposition 5. First, for regime 2 recall that the regulatory constraint becomes binding

[25] Note, however, that this comparative static does not hold globally.

[26] As will be shown below, this need not be the case, i.e., depending on acquisition costs k, there are also cases where $\mu_1^* < \overline{\mu}$.

just when $\tau = T_2^*$. From there on, the constrained optimal choice of implemented diligence effort will be as low as is feasible (so as to satisfy both the regulatory constraint and the agent's acquisition constraint). That is, when regulation binds in regime 2 we always have that $\mu^*(\tau) = \underline{\mu}(\tau)$, which - as we know - is strictly increasing in τ. When we are initially in regime 3, however, where $b_0 > 0$, there is an additional positive effect of regulation on diligence. Intuitively, while an up-front bonus $b_0 > 0$ that is paid in the absence of regulation does not generate incentives for the agent to exert higher diligence, this is the case for any other contingent bonus that is paid with at least some delay. While this mandatory delay of compensation certainly increases the level of compensation costs, it decreases the *marginal* cost of inducing diligence in this case (see proof of Proposition 5), leading to an increase in equilibrium diligence. The case where initially regime 3 applies is illustrated in Figure 5. Due to the presence of an up-front bonus without regulation, regulation is effective even for small τ.

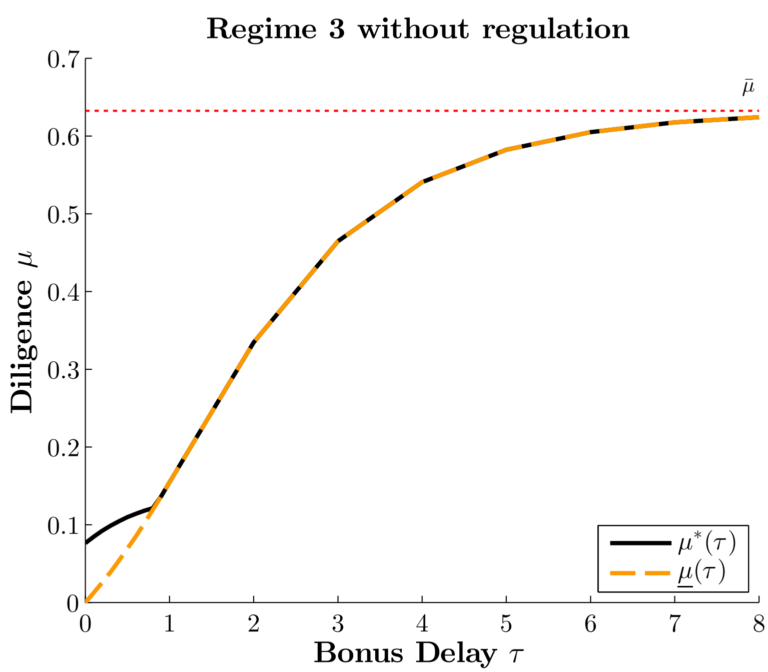

Figure 5. This graph plots the equilibrium diligence μ^* as a function of the mandatory delay of compensation τ for the case when regime 3 obtains without regulation. Since an upfront bonus is paid without regulation, regulation has an immediate positive effect on equilibrium diligence. Once the constraint $b_T \geq 0$ binds, the implementation constraint $\underline{\mu}(\tau)$ determines the equilibrium diligence level (and regime 2 obtains). The revenue function satisfies $\pi(\mu) = 1.5\mu$. The remaining parameters are as in Figure 4.

Discussion. Combining Propositions 4 and 5, we can immediately make the following general observations on the maximum level of diligence that can be implemented via regulation (or lack thereof). These observations will be used subsequently to derive, in a particular case, the optimal deferral period (see Section 5).

When $k \geq \underline{k}^*$ (regimes 2 and 3 without regulation) we have shown that the imposition of mandatory deferral increases diligence and this holds as well for any further increase in the minimum deferral period. Note, however, that once the regulatory constraint becomes binding, bank profits are strictly decreasing in τ and will be strictly negative when τ exceeds some cutoff value. If regulation has to satisfy also the bank's participation constraint, i.e., when it must ensure at least zero profits from this line of business, this imposes an upper bound on τ. Then, when $k \geq \underline{k}^*$, the deferral period that maximizes diligence, while ensuring at least zero profits, is given by the respective threshold value, which we denote by $\overline{\tau}$.[27]

Recall next that when originally the acquisition constraint is slack with $k < \underline{k}^*$ so that we are in regime 1, there are two cases to consider, depending on how the unconstrained equilibrium choice $\mu^*(\tau = 0)$ in regime 1 compares with the (limit) threshold $\overline{\mu}$. Note here that, as the acquisition constraint is slack in this regime, μ^* does not depend on k. On the other hand, recall how acquisition effort costs k affect $\overline{\mu}$, i.e., the minimum level of diligence that a bank has to implement when the bonus is (almost) infinitely delayed (while it must still satisfy the agent's acquisition incentive constraint): $\overline{\mu}$ is strictly increasing in k, and we also know that $\overline{\mu}$ goes to zero when k approaches zero. So whether $\mu^*(\tau = 0)$ in regime 1 is larger or small than the maximum diligence that can be achieved with mandatory deferral depends also on the acquisition costs k. Taken together, we obtain the following clear-cut results on whether mandatory deferral can increase equilibrium diligence.

Proposition 6 *There exists a cutoff on the costs of acquisition effort, $\widehat{k}$ so that for low values of $k < \widehat{k}$ the highest equilibrium diligence $\mu^*(\tau)$ is achieved when no binding deferral regulation at all is imposed. Instead, for all $k > \widehat{k}$ regulation can lead to higher*

[27] While this paper only considers binary acquisition effort, we conjecture that, when allowing for acquisition effort to be a continuous choice (cf. also Appendix B), the bank responds to an increasingly restrictive regulation (higher τ) by gradually reducing the level of customer or deal acquisition that it wants to induce. This may then also require to reduce the optimally implemented level of diligence effort, given the complementarity of the two tasks. For sufficiently high τ, we thus expect $\mu^*(\tau)$ to be decreasing in τ, such that, also in this case, $\mu^*(\tau)$ is non-monotonic.

diligence. In particular, equilibrium diligence would then be highest when the minimum deferral period τ is made as high as possible, i.e., $\tau = \overline{\tau}$ in case the bank's zero-profit constraint must be satisfied.

Proof. See Appendix A.

5 Optimal Mandatory Deferral Regulation

The main part of our paper focused on the *positive* question when the particular policy of mandatory deferral leads to lower diligence and when it can instead induce higher equilibrium diligence. For this positive analysis, we neither needed to assert that the tool of mandatory deferral is optimal among all feasible regulatory policies nor that the regulator actually chooses the optimal deferral time within the narrow class of deferral regulations. However, we implicitly assumed that the regulator's underlying rationale to regulate was driven by the motivation to increase equilibrium diligence, which allowed us to identify situations in which binding deferral regulation must "backfire" (see Proposition 6).

We now turn to the normative question regarding an optimal deferral period for one stylized application (see Section 5.1) that derives from primitives the revenue function of the bank $\pi(\mu)$ as well as defines a welfare function Ω, the objective function of the regulator. The bank's objective function and the one of the regulator differ due to a negative externality imposed by the bad event. In the subsequent Section 5.2, we then discuss the general insights for the design of deferral regulation as a function of economic primitives. We note that our normative analysis still leaves open the question whether a broader class of regulatory tools could achieve more efficient outcomes. In particular, if the regulator could ex post impose a sufficiently large penalty in case the bad event occurs, the first-best outcome could always be trivially achieved. Thus, for deferral regulation to be optimal in a broader class of regulatory policies, there must be reasons, such as limited liability of banks or ex post enforceability of such a penalty, that make it impossible to have the bank internalize the cost in such a simple way.

5.1 Application: Loan Defaults with Externalities

We consider a single consumer loan or mortgage that the bank employee generates with probability a. Through exerting (diligence) effort μ the agent can decrease the likelihood with which a loan subsequently defaults. This application follows Hartman-Glaser et al. (2012). In this example there may be various reasons for why the privately induced level of diligence by the bank is not welfare optimal. Even when there is no moral hazard problem in the contractual relationship between the bank and the household that takes out a loan, both parties should fail to internalize negative effects on other parties that would arise from default. Campbell et al. (2011) document such negative spillover from foreclosures on local house prices, and these externalities may not only be pecuniary.

Suppose that the bank makes a consumer loan of size 1, designed as a perpetuity with flow payment F. We stipulate for simplicity that the bank acts as a monopolist and can, through interest payments, extract the full consumer surplus from any given loan. Further, we assume that upon default, the "bad event", the asset becomes worthless, while there are no private costs of bankruptcy for the borrower. These stylized assumptions allow us to write the private revenue of the bank (which includes consumer surplus) as:

$$\pi(\mu) = F\left[\frac{1}{r+\lambda_H} + \mu\left(\frac{1}{r+\lambda_L} - \frac{1}{r+\lambda_H}\right)\right] - 1. \tag{17}$$

We can now envisage that default leads to a non-internalized social loss of size $X > 0$. Then, the expected negative externality for a given level of diligence and given a discount rate r for X is

$$\omega(\mu) = X\left[\frac{\lambda_H}{r+\lambda_H} - \mu\left(\frac{\lambda_H}{r+\lambda_H} - \frac{\lambda_L}{r+\lambda_L}\right)\right]. \tag{18}$$

When no acquisition effort is chosen, so that $a = a_l = 0$, then obviously total welfare can be written as $\Omega = \Omega_l = 0$. When there is acquisition, so that $a = a_h = 1$, we can write welfare as $\Omega_h(\tau) = \Pi + V_A - \omega$, i.e., as the sum of the bank's and the agent's expected payoff minus the expected externality. Note that we have here suppressed the dependency on $\mu^*(\tau)$ but also on the full compensation contract that is thereby induced. As the compensation contract is uniquely determined, we can make this explicit as follows:

$$\Omega_h(\tau) = \Pi(\mu^*(\tau), \tau) + V_A(\mu^*(\tau), \tau) - \omega(\mu^*(\tau)). \tag{19}$$

Recall now from Proposition 6 that there is a level $\tau = \overline{\tau}$ from which on the bank would stop operations. Consequently, we define

$$\Omega_h^*(\tau) = \max_{\tau \leq \overline{\tau}} \Omega_h(\tau).$$

In what follows, we always stipulate that externalities are sufficiently small to ensure that $\Omega_h^*(\tau) > \Omega_l = 0$. Otherwise, it would be (weakly) optimal to close down this particular line of the bank's business, which could be ensured also by stipulating that compensation must be deferred beyond time $\overline{\tau}$. Consider, thus, a regulator's problem to choose a minimum deferral period τ so as to maximize welfare under acquisition, $\Omega_h(\tau)$.

Proposition 7 *Binding regulation via mandatory deferral is only optimal if both $k > \widehat{k}$ and externalities X are sufficiently high. Then, the optimal deferral time is strictly increasing in X.*

Proof. See Appendix A.

The intuition behind these two conditions is as follows. First, if $k < \widehat{k}$, we can already rely on Proposition 6 to conclude that deferral regulation cannot improve welfare since regulation can only lead to lower diligence (and the externality would, if anything, call for higher diligence). For $k > \widehat{k}$ instead a higher level of diligence can be achieved through mandating deferral of compensation. However, it is still unclear whether this increase in diligence is welfare-enhancing since deferral also has a first-order effect on the deadweight loss D.

This is why the size of the externality X matters. Take the benchmark without externalities, so that $X = 0$ and, thus, $\omega(\mu) = 0$ for all μ. We show in the proof of Proposition 7 that, despite the agency conflict between the bank and the agent, the unregulated choice maximizes welfare (this is in particular obvious when the acquisition constraint binds, i.e., $V_A = 0$). Hence, if there is scope for welfare improving mandatory deferred compensation, in our model this can *only* be due to an externality. Moreover, we can show that in order to achieve even only a marginal increase in diligence, the resulting increase in deadweight loss D is non-marginal. Consequently, when X is small mandatory deferral is never optimal regardless of k. Finally, once the externality is sufficiently high to warrant intervention, it is fairly intuitive that the optimum deferral time (which increases diligence) becomes larger.

5.2 Normative Implications for Regulation Design

The insights from the just discussed example should be applicable to other (and less stylized) settings. According to Proposition 7, mandatory deferral can only be optimal from a welfare perspective when acquisition is sufficiently important and when externalities are sufficiently large as well. In particular the empirical role of acquisition cost k and the respective threshold $\widehat{k}$ are worthwhile to discuss. When k is relatively high, this may suggest that deal or customer acquisition is a main part of the agent's work. Propositions 6 and 7 then suggest that imposing mandatory deferral backfires when the respective agent acts, in terms of the importance of the respective tasks for compensation, more like a "bureaucrat." (Then, in case of retail financial products, these would be rather "bought" than "sold.") Alternatively, variation in k could capture factors that make it more or less difficult to generate new customers and deal opportunities. When we stipulate that more competition raises k, then Propositions 6 and 7 would suggest that regulation of deferred incentive pay could be (more) beneficial when competition intensifies.

Alternatively, we can consider variations in the threshold for acquisition cost $\widehat{k}$. Recall that $\widehat{k}$ depends on a comparison with the equilibrium outcome in the absence of regulation, $\mu^*(\tau = 0)$: The higher $\mu^*(\tau = 0)$, the higher is the threshold for k, so that the range of parameters increases for which any binding mandatory deferral leads to lower diligence. Depending on the application, it is often possible to identify observable determinants that govern the sensitivity of the profit function $\pi(\mu)$ to diligence (and hence capture variation in the unregulated effort choice). For instance, the bank arguably cares more about diligence when it keeps a larger fraction of loans on its own books. As a result, deferral is more likely to be ineffective when the degree of securitization is low. Similarly, stricter legal enforcement of liability in the case of misselling or unsuitable advice should raise incentives for diligence in the unregulated benchmark (without deferral compensation).[28] The latter example points to interesting spillover effects between different types of regulatory tools. If regulation already targets incentives directly such as via enforcement of liability (through $\pi(\mu)$), then additional regulation via mandatory deferral in compensation is more likely to backfire.

6 Concluding Remarks

The first part of this paper presents a characterization of the optimal compensation when a principal wants to induce a given level of acquisition and diligence effort. The first effort determines the likelihood with which a "deal" arises in the first place, while diligence reduces the likelihood that such a deal generates a critical, bad event. Key

[28]Note that such changes would affect the function $\pi(\mu)$, but not the primitives used for our characterization, such as Propositions 1 and 3 (i.e., k, Δ_r, and Δ_λ).

applications, as discussed, are loans as well as the sale of long-term financial or insurance products to retail customers, which - when unsuitable - may lead to cancellations or even liability and reputational problems. Moreover, the principal who designs the optimal compensation contract, i.e., the bank in our applications, may not fully internalize all effects that arise from such a "bad event." This may hold as third parties are affected, but also when limited observability and commitment as well as naiveté preclude efficient contracting between the bank and its contractual party to the deal, such as a household taking out a loan or signing up to a savings plan. In these cases, there may be scope for regulation that induces a higher level of diligence than what would otherwise arise in equilibrium. We analyze whether this is indeed achieved through a policy of mandating a longer deferral of bonus payments.

While such a mandatory deferral indeed makes available more information until a contingent payment is made, we show that it may not induce higher but rather lower diligence effort. One key insight is that as it distorts the bank's optimal use of contractual instruments, it raises not only the overall compensation costs for a given level of diligence but may also raise the marginal compensation costs for inducing higher diligence. As a consequence, the bank may react to the regulation by optimally inducing a lower rather than a higher level of diligence. However, we also identify positive effects from a mandatory deferral. Notably, when acquisition requires sufficiently high incentives, then without regulation this possibly leads to a large up-front bonus that is not made contingent on subsequent performance of a deal. Intuitively, in this case mandatory deferral can ensure that also this component of pay provides incentives to exert diligence rather than only acquisition or deal-making effort. But also the bank's own incentives to elicit diligence effort are key in predicting how it will respond to regulation.

Still, note that in this paper we analyze only the implications of a particular regulatory proposal, namely to impose a minimum mandatory deferral time, rather than asking the broader question of optimal regulation. Future work could turn to the question when deferral regulation should optimally be part of unified financial regulation which takes into account the interdependencies of various regulatory tools.

References

Acharya, V.V. and P.F. Volpin. 2010. Corporate Governance Externalities. *Review of Finance* 14, 1–33.

Bebchuck, L.A., and J.M. Fried. 2004. Pay without performance: the unfulfilled promise of executive compensation. Cambridge : Harvard University Press.

Bebchuck, L.A., and J.M. Fried. 2010. Paying for Long-Term Performance. *University of Pennsylvania Law Review* 158, 1915–1960.

Bénabou, R. and J. Tirole. 2013. Bonus Culture: Competitive Pay, Screening, and Multitasking. Mimeo, NBER Working Paper 18936.

Biais, B., T. Mariotti, J.-C. Rochet, and S. Villeneuve. 2010. Large Risks, Limited Liability, and Dynamic Moral Hazard. *Econometrica* 78, 73–118.

Bolton, P., H. Mehran, and J. Shapiro. 2010. Executive Compensation and Risk Taking. Mimeo, Columbia University.

Brisley, N. 2006. Executive Stock Options: Early Exercise Provisions and Risk-Taking Incentives. *Journal of Finance* 61, 2487–2509.

Campbell, J.Y., S. Giglio, and P. Pathak. 2011. Forced Sales and House Prices. *American Economic Review* 101, 2108–2131.

Chaigneau, P. 2012. The Optimal Timing of CEO Compensation. Mimeo, HEC Montreal.

DeMarzo, P., and D. Duffie. 1999. A Liquidity-Based Model of Security Design. *Econometrica* 67, 65–99.

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

Dicks, D.L. 2012. Executive Compensation and the Role of Corporate Governance Regulation. *Review of Financial Studies* 25, 1971–2004.

Edmans, A., X. Gabaix, T. Sadzik, and Y. Sannikov. 2012. Dynamic CEO Compensation. *Journal of Finance* 67, 1603–1647.

Fahlenbrach, R., and R. Stulz. 2011. Bank CEO Incentives and the Credit Crisis. *Journal of Financial Economics* 99, 11–26.

Gromb, D., and D. Martimort. 2007, Collusion and the Organization of Delegated Expertise. *Journal of Economic Theory* 137, 271–299.

Hartman-Glaser, B., T. Piskorski, and A. Tchistyi. 2012. Optimal Securitization with Moral Hazard. *Journal of Financial Economics* 104, 186–202.

Heider, F., and R. Inderst. 2012. Loan Prospecting. *Review of Financial Studies* 25, 2381–2415.

Inderst, R., and S. Pfeil. 2013. Securitization and Compensation in Financial Institutions. *Review of Finance* 17, 1323–1364.

Kuhnen, C., and J. Zwiebel. 2009. Executive Pay, Hidden Compensation and Managerial Entrenchment. Mimeo, Stanford University.

Laux, V. 2012. Stock Option Vesting Conditions, CEO Turnover, and Myopic Investment. *Journal of Financial Economics* 106, 513–526.

Levitt, S.D., and C.M. Snyder. 1997. Is No News Bad News? Information Transmission and the Role of 'Early Warning'. *Rand Journal of Economics* 28, 641–661.

Malamud, S., H. Rui, and A. Whinston. 2013. Optimal Incentives and Securitization of Defaultable Assets. *Journal of Financial Economics* 107, 111–135.

Peng, L., and A. Roell. 2011. Managerial Incentives and Stock Price Manipulation. Mimeo, Baruch College.

Rogerson, W.P. 1997. Intertemporal Cost Allocation and Managerial Investment Incentives: A Theory Explaining the Use of Economic Value Added as a Performance Measure. *Journal of Political Economy* 105, 770–795.

Thanassoulis, J.E. 2012. The Case for Intervening in Bankers' Pay. *Journal of Finance* 67, 849–895.

Appendix A: Omitted Proofs

Proof of Propositions 1 and 3. It is convenient to restate the full program, where we take as given that the bank wants to implement μ as well as $a = a_h = 1$:

$$\min_{b_i,T_i} \left\{ \sum\nolimits_i b_i e^{-r_P T_i} \left[\mu e^{-(\lambda_H - \Delta_\lambda)T_i} + (1-\mu) e^{-\lambda_H T_i} \right] \right\}$$

$$s.t.$$

$$\sum\nolimits_i b_i e^{-r_A T_i} \left(e^{-(\lambda_H - \Delta_\lambda)T_i} - e^{-\lambda_H T_i} \right) = c'(\mu), \tag{20}$$

$$\sum\nolimits_i b_i e^{-r_A T_i} \left[\mu e^{-(\lambda_H - \Delta_\lambda)T_i} + (1-\mu) e^{-\lambda_H T_i} \right] - c(\mu) \geq k, \tag{21}$$

$$T_i \geq \tau,$$

$$b_i \geq 0.$$

When no regulatory constraint is imposed we have $\tau = 0$, while, else, $\tau > 0$. Define κ_μ (for (20)), κ_a (for (21)), κ_{T_i} (for each T_i) and κ_{b_i} (for each b_i) as the respective Lagrange multipliers of the problem.

Clearly, as we can add up all payments b_i made at the same time T_i, the constraint $T_i \geq \tau$ binds at most once and we denote the associated payment at $T_0 = \tau$ by b_τ. Hence, $\kappa_{T_i} = 0$ for all $i \geq 1$.

Now, the first-order condition with respect to b_i is given by

$$\left(e^{\Delta_r T_i} - \kappa_a \right) \left[1 + \mu \left(e^{\Delta_\lambda T_i} - 1 \right) \right] - \kappa_\mu \left(e^{\Delta_\lambda T_i} - 1 \right) - e^{(r_A + \lambda_H)T_i} \kappa_{b_i} = 0.$$

This holds for any b_i. For $i = 0$, i.e., for the payment at τ (b_τ), we can rewrite the first-order condition to obtain[29]

$$\kappa_a = e^{\Delta_r \tau} - \frac{e^{(r_A + \lambda)\tau} \kappa_{b_\tau} + \kappa_\mu \left(e^{\Delta_\lambda \tau} - 1 \right)}{1 + \mu \left(e^{\Delta_\lambda \tau} - 1 \right)}. \tag{22}$$

For $i \geq 1$ and $\kappa_{b_i} = 0$ we can thus write

$$\kappa_\mu = \left(1 + \left(e^{\Delta_\lambda T_i} - 1 \right) \mu \right) \frac{e^{(r_A+\lambda)\tau} \kappa_{b_\tau} + \left(e^{\Delta_r T_i} - e^{\Delta_r \tau} \right) (1-\mu) + \left(e^{\Delta_\lambda \tau + \Delta_r T_i} - e^{(\Delta_r + \Delta_\lambda)\tau} \right) \mu}{e^{\Delta_\lambda T_i} - e^{\Delta_\lambda \tau}}. \tag{23}$$

Consider next the first-order condition with respect to $T_i > \tau$. Note that if $\kappa_{b_i} > 0$, i.e., $b_i = 0$, the first-order condition with respect to T_i is trivially satisfied. When $\kappa_{b_i} = 0$,

[29] In the special case where $\tau = 0$ we can simplify this expression to get $\kappa_a = 1 - \kappa_{b_0}$, which implies, together with $\kappa_a \geq 0$ and $\kappa_{b_0} \geq 0$, that $0 \leq \kappa_{b_0} \leq 1$.

substituting from (22) and (23) for κ_a and κ_μ, any $T_i > \tau$ must satisfy

$$\Delta_r \left(1 - e^{-\Delta_\lambda (T_i - \tau)}\right)\left(1 + \mu\left(e^{\Delta_\lambda T_i} - 1\right)\right) - \Delta_\lambda \left(1 - e^{-\Delta_r (T_i - \tau)}\right)\left(1 + \mu\left(e^{\Delta_\lambda \tau} - 1\right)\right) \quad (24)$$
$$= \Delta_\lambda e^{(r_A + \lambda)\tau - \Delta_r T_i} \kappa_{b_\tau}.$$

Lemma A1. *Consider equation (24) with the restriction to $\kappa_{b_\tau} \geq 0$. If $\frac{\Delta_r}{\Delta_\lambda} < \frac{1-\mu\left(1+e^{\Delta_\lambda \tau}\right)}{1+\mu\left(e^{\Delta_\lambda \tau}-1\right)}$, then there exists a unique $T_i = T > \tau$ solving equation (24). If $\frac{\Delta_r}{\Delta_\lambda} \geq \frac{1-\mu\left(1+e^{\Delta_\lambda \tau}\right)}{1+\mu\left(e^{\Delta_\lambda \tau}-1\right)}$, then a solution exists only if $\kappa_{b_\tau} > 0$ and it is again unique.*

Proof. Consider the following functions appearing on the left- and right-hand-side of (24) respectively:

$$f(T,\tau) = \Delta_r \left(1 - e^{-\Delta_\lambda (T - \tau)}\right)\left(1 + \mu\left(e^{\Delta_\lambda T} - 1\right)\right) - \Delta_\lambda \left(1 - e^{-\Delta_r (T - \tau)}\right)\left(1 + \mu\left(e^{\Delta_\lambda \tau} - 1\right)\right),$$
$$g(T,\tau) = \Delta_\lambda e^{(r_A + \lambda)\tau - \Delta_r T} \kappa_{b_\tau}.$$

For $g(T,\tau)$ we have the following simple properties: If $\kappa_{b_\tau} = 0$, then $g(T,\tau) = 0$ for all T, otherwise it holds that $g(T,\tau) > 0$ and $\partial g(T,\tau)/\partial T < 0$ for all $T \geq \tau$. Next, the function $f(T,\tau)$ satisfies

$$f(\tau,\tau) = \left.\frac{\partial f(T,\tau)}{\partial T}\right|_{T=\tau} = 0,$$
$$\left.\frac{\partial^2 f(T,\tau)}{\partial T^2}\right|_{T=\tau} = \Delta_\lambda \Delta_r \left[\Delta_\lambda \left(\mu\left(1 + e^{\Delta_\lambda \tau}\right) - 1\right) + \Delta_r \left(\mu\left(e^{\Delta_\lambda \tau} - 1\right) + 1\right)\right],$$
$$\lim_{T\to\infty} f(T,\tau) = \infty.$$

Depending on whether $\partial^2 f(T,\tau)/\partial T^2$ is positive or negative we now distinguish two cases.

Case $\frac{\Delta_r}{\Delta_\lambda} \geq \frac{1-\mu\left(1+e^{\Delta_\lambda \tau}\right)}{1+\mu\left(e^{\Delta_\lambda \tau}-1\right)}$: In this case f is convex at $T = \tau$. We will show, that this implies that f is increasing for all $T > \tau$, such that from the properties of g together with $\lim_{T\to\infty} f(T,\tau) = \infty$ there exists a unique solution $T > \tau$ to (24) if and only if $\kappa_{b_\tau} > 0$. Note that

$$\frac{\partial f(T,\tau)}{\partial T} = \Delta_\lambda \Delta_r \left[e^{-\Delta_\lambda (T-\tau)} + \mu\left(e^{\Delta_\lambda T} - e^{-\Delta_\lambda (T-\tau)}\right) - e^{-\Delta_r (T-\tau)}\left(1 + \mu\left(e^{\Delta_\lambda \tau} - 1\right)\right)\right],$$

such that the sign of $\partial f\left(T,\tau\right)/\partial T$ is determined by the term in square brackets, which we denote by $H(T)$. Using $\frac{\Delta_r}{\Delta_\lambda} \geq \frac{1-\mu\left(1+e^{\Delta_\lambda \tau}\right)}{1+\mu\left(e^{\Delta_\lambda \tau}-1\right)}$ it holds that

$$\begin{aligned} H(T) &= e^{-\Delta_\lambda (T-\tau)} + \mu\left[e^{\Delta_\lambda T} - e^{-\Delta_\lambda (T-\tau)}\right] - e^{-\Delta_r (T-\tau)}\left(1+\mu\left(e^{\Delta_\lambda \tau}-1\right)\right) \\ &\geq e^{-\Delta_\lambda (T-\tau)}\left[(1-\mu) + \underbrace{\mu e^{\Delta_\lambda (2T-\tau)} - e^{\Delta_\lambda \frac{2\mu e^{\Delta_\lambda \tau}}{1+\mu\left(e^{\Delta_\lambda \tau}-1\right)}(T-\tau)}\left(1+\mu\left(e^{\Delta_\lambda \tau}-1\right)\right)}_{=:h(T)}\right]. \end{aligned}$$

The result that $\partial f\left(T,\tau\right)/\partial T > 0$ for all $T > \tau$ then follows from $h(\tau) = -\left(1-\mu\right)$ together with

$$h'(T) = 2\Delta_\lambda \mu e^{\Delta_\lambda \tau}\left(e^{2\Delta_\lambda (T-\tau)} - e^{\frac{\mu e^{\Delta_\lambda \tau}}{(1-\mu)+\mu e^{\Delta_\lambda \tau}} 2\Delta_\lambda (T-\tau)}\right) > 0.$$

Case $\frac{\Delta_r}{\Delta_\lambda} < \frac{1-\mu\left(1+e^{\Delta_\lambda \tau}\right)}{1+\mu\left(e^{\Delta_\lambda \tau}-1\right)}$: Here, existence follows trivially from $f(\tau,\tau) = \partial f(\tau,\tau)/\partial T = 0$, together with $\partial^2 f(\tau,\tau)/\partial T^2 < 0$ and $\lim_{T\to\infty} f\left(T,\tau\right) = \infty$, together with the properties of $g(T,\tau)$. What remains to be shown is uniqueness. We argue to a contradiction. Assume thus that $f(T,\tau)$ and $g(T,\tau)$ intersect more than once. Then, as $f(\tau,\tau) = \partial f(\tau,\tau)/\partial T = 0$ and $\partial^2 f(\tau,\tau)/\partial T^2 < 0$, there must exist a $\widetilde{T} > 0$ where $f(T,\tau)$ changes its curvature from convex to concave, i.e., $\partial^2 f(\widetilde{T},\tau)/\partial T^2 = 0$ and $\partial^3 f(\widetilde{T},\tau)/\partial T^3 < 0$. So, from

$$\begin{aligned} &\frac{\partial^2 f(T,\tau)}{\partial T^2} \\ &= \Delta_\lambda \Delta_r\left[-\Delta_\lambda e^{-\Delta_\lambda (T-\tau)} + \mu \Delta_\lambda\left[e^{-\Delta_\lambda (T-\tau)} + e^{\Delta_\lambda T}\right] + \Delta_r e^{-\Delta_r (T-\tau)}\left(1+\mu\left(e^{\Delta_\lambda \tau}-1\right)\right)\right] \end{aligned}$$

it must hold that

$$\Delta_r e^{-\Delta_r\left(\widetilde{T}-\tau\right)}\left(1+\mu\left(e^{\Delta_\lambda \tau}-1\right)\right) = \Delta_\lambda e^{-\Delta_\lambda\left(\widetilde{T}-\tau\right)} - \mu \Delta_\lambda\left[e^{-\Delta_\lambda\left(\widetilde{T}-\tau\right)} + e^{\Delta_\lambda \widetilde{T}}\right].$$

Substituting in

$$\begin{aligned} &\frac{\partial^3 f(T,\tau)}{\partial T^3} \\ &= \Delta_\lambda \Delta_r\left[\Delta_\lambda e^{-\Delta_\lambda (T-\tau)} + \mu \Delta_\lambda\left[e^{\Delta_\lambda T} - e^{-\Delta_\lambda (T-\tau)}\right] - \Delta_r e^{-\Delta_r (T-\tau)}\left(1+\mu\left(e^{\Delta_\lambda \tau}-1\right)\right)\right] \end{aligned}$$

gives

$$\begin{aligned} &\frac{\partial^3 f(\widetilde{T},\tau)}{\partial T^3} \\ &= \Delta_\lambda \Delta_r e^{-\Delta_\lambda\left(\widetilde{T}-\tau\right)}\left[(\Delta_\lambda - \Delta_r) + \mu \Delta_\lambda\left[e^{\Delta_\lambda\left(2\widetilde{T}-\tau\right)} - 1\right] + \mu \Delta_r\left[1 + e^{\Delta_\lambda\left(2\widetilde{T}-\tau\right)}\right]\right] > 0, \end{aligned}$$

where we have used that $\Delta_r < \Delta_\lambda$, which follows from $\frac{\Delta_r}{\Delta_\lambda} < \frac{1-\mu\left(1+e^{\Delta_\lambda \tau}\right)}{1+\mu\left(e^{\Delta_\lambda \tau}-1\right)}$, contradiction. **Q.E.D.**

From Lemma A1 we know that there are at most two payments one at τ and/or one at $T > \tau$. Using this result, we will now first characterize the optimal contract for the case where $\tau = 0$, then, second, the optimal contract for $\tau > 0$. For $\tau = 0$, consider three different cases, corresponding to different values of κ_{b_0}.

Case $\tau = 0$, $\kappa_{b_0} = 0$. Then, the from Lemma A1 unique solution $T > 0$ must satisfy (11). The associated payment b_T then follows from (20) and is given by (5). Finally, (22) together with $\kappa_{b_0} = 0$ then imply that $\kappa_a = 1$. Hence, (21) must hold with equality so that b_0 is given by (7). Finally, this case applies if and only if $\frac{\Delta_r}{\Delta_\lambda} < 1 - 2\mu$ and $k \geq \overline{k}$, where $\overline{k}$ in (13) is obtained from setting $b_0 = 0$ in (7).

Case $\tau = 0$, $\kappa_{b_0} = 1$. Then, (22) implies $\kappa_a = 0$, such that the constraint (21) is slack. The first-order condition (24) simplifies to

$$\Delta_r \left(1 - e^{-\Delta_\lambda T}\right) + \Delta_r \mu \left(e^{\Delta_\lambda T} + e^{-\Delta_\lambda T} - 2\right) - \Delta_\lambda = 0, \tag{25}$$

which has a unique solution given by (9), while again b_T is given by (5) and now $b_0 = 0$. This case applies if and only if $k < \underline{k}$, where $\underline{k}$ as given in (12) is obtained from the slack constraint (21). We finally show that $\underline{k} < \overline{k}$, which is equivalent to showing that $T = T'$ solving (25) is larger than $T = T''$ solving (11). This follows as the left-hand-side in (25) is increasing in T and as, when evaluated at $T = T''$ becomes $-\Delta_\lambda e^{-\Delta_r T''} < 0$.

Case $\tau = 0$, $\kappa_{b_0} \in (0, 1)$. Then, $b_0 = 0$ and, from (22), the constraint (21) binds. We now obtain from (24), (20), and (21) explicit solutions for T and b_T which are given by (10) and

$$b_T = \left[k + c(\mu) - c'(\mu)\mu\right] \left(1 + \frac{c'(\mu)}{k + c(\mu) - c'(\mu)\mu}\right)^{\frac{r_A + \lambda_H}{\Delta_\lambda}},$$

respectively. By the preceding characterization, this case applies whenever $k \geq \underline{k}$ and $\frac{\Delta_r}{\Delta_\lambda} \geq 1 - 2\mu$, and, for $\frac{\Delta_r}{\Delta_\lambda} < 1 - 2\mu$, if $\underline{k} \leq k \leq \overline{k}$.

This completes the proof of Proposition 1. Continuing with the proof of Proposition 3, consider now $\tau > 0$. We will distinguish two different cases, corresponding to whether the acquisition constraint binds or not.

Case $\tau > 0$ and slack acquisition constraint. From the preceding observations we have that the acquisition constraint is slack for $\tau = 0$ if $k < \underline{k}$. In this case there is a single bonus paid at $T_1 > 0$ as given by (9). The characterization for $\tau > 0$ then follows from the fact that implementation costs are monotonically increasing for $T > T_1$ as was shown above (cf. the left-hand-side in (25)). The unique payment is then given by (5) with $T = T_1$ for $\tau \leq T_1$ and $T = \tau$ for $\tau > T_1$.[30]

Case $\tau > 0$ and binding acquisition constraint. Take now the case where $k \geq \underline{k}$. When (15) is violated and, hence, from Lemma A1 there is only a single payment, the two binding constraints (20) and (21) imply that this occurs at $T = T_2$ as defined in (10) and is given by (5). Hence, from Proposition 2, the respective diligence level can only be implemented as long as the regulatory constraint does not bind, i.e., as long as $\tau \leq T_2$.

Now, when (15) holds there can be two positive payments, which are determined from the binding constraints (20) and (21):

$$b_\tau = \frac{e^{(r_A+\lambda_H)\tau}\left(e^{\Delta_\lambda T}-1\right)}{\left(e^{\Delta_\lambda T}-e^{\Delta_\lambda \tau}\right)}\left[k+c\left(\mu\right)-\left[\mu+\frac{1}{\left(e^{\Delta_\lambda T}-1\right)}\right]c'\left(\mu\right)\right],$$

$$b_T = -\frac{e^{(r_A+\lambda_H)T}\left(e^{\Delta_\lambda \tau}-1\right)}{\left(e^{\Delta_\lambda T}-e^{\Delta_\lambda \tau}\right)}\left[k+c\left(\mu\right)-\left[\mu+\frac{1}{\left(e^{\Delta_\lambda \tau}-1\right)}\right]c'\left(\mu\right)\right].$$

From these expressions we directly have that $b_T > 0$ as long as $\tau < T_2$, i.e., as long as μ is implementable. For $b_\tau > 0$, we require that $T_3(\tau)$ solving (16) satisfies $T_3(\tau) > T_2$. Hence, there are two positive payments b_τ and b_T, if and only if μ is implementable, (15) holds and $k > \overline{k}(\tau) := \left[\mu + \frac{1}{\left(e^{\Delta_\lambda T_3(\tau)}-1\right)}\right]c'\left(\mu\right) - c(\mu)$. In all other cases there is a single payment at $T = T_2$.

When a positive payment at τ and $T = T_3(\tau)$ is made, i.e., $\kappa_{b_\tau} = 0$ and $\kappa_{b_T} = 0$, we have

$$sgn\left(\frac{dT_3}{d\tau}\right) = sgn\left(-\frac{\frac{\partial f}{\partial \tau}}{\frac{\partial f}{\partial T}}\right) = -sgn\left(\frac{\partial f}{\partial \tau}\right),$$

where the second equality follows from the fact that at $T = T_3(\tau)$ it holds that $\frac{\partial f}{\partial T} > 0$ (cf. proof of Lemma A1). The relevant part of $\frac{\partial f}{\partial \tau}$ then is

$$sgn\left(\frac{\partial f}{\partial \tau}\right) = sgn\left(\Delta_r\left(1-\mu\right)\left(e^{\Delta_\lambda(T-\tau)}-e^{\Delta_r(T-\tau)}\right)+\left(\Delta_r+\Delta_\lambda\right)\mu e^{\Delta_\lambda T}\left(1-e^{\Delta_r(T-\tau)}\right)\right). \tag{26}$$

[30]Note that for $k \leq \mu c'(\mu) - c(\mu)$ the acquisition constraint is slack for all τ. If, however, $k > \mu c'(\mu) - c(\mu)$, then (21) remains slack only as long as $\tau \leq T_2$, i.e., as long as μ remains implementable (cf. Proposition 2).

Since in the relevant up-front payment region $\Delta_\lambda > \Delta_r$, the first term is positive and the second term is negative. Using the first-order condition for T i.e., $f(T_3, \tau) = 0$, solving $f(T_3, \tau)$ for μ (see equation 24) and substituting this into (26) yields

$$sgn\left(\frac{\partial f}{\partial \tau}\right) = sgn\left(e^{\Delta_\lambda T}\frac{J_1}{e^{\Delta_\lambda \tau}J_2 + J_3}\right) \text{ where:}$$

$$J_1 = {\Delta_r}^2\left(e^{\Delta_\lambda(T-\tau)} + e^{-\Delta_\lambda(T-\tau)} - 2\right) - \Delta_\lambda^2\left(e^{\Delta_r(T-\tau)} + e^{-\Delta_r(T-\tau)} - 2\right) > 0,$$

$$J_2 = \Delta_r\left(e^{\Delta_\lambda(T-\tau)} - 1\right) - \Delta_\lambda\left(1 - e^{-\Delta_r(T-\tau)}\right) > 0,$$

$$J_3 = \Delta_\lambda\left(1 - e^{-\Delta_r(T-\tau)}\right) - \Delta_r\left(1 - e^{-\Delta_\lambda(T-\tau)}\right) > 0.$$

Noting that, from $\Delta_\lambda > \Delta_r$ and the convexity of the exponential function, each of the terms J_1 to J_3 is positive, we find that $\frac{\partial f}{\partial \tau} > 0$, implying $\frac{dT_3}{d\tau} < 0$. **Q.E.D.**

Proof of Corollary 2. When T is determined by (9), we have

$$\frac{dT}{d\mu} = -\frac{e^{-\Delta_\lambda T}\left(e^{\Delta_\lambda T} - 1\right)^2}{\Delta_\lambda\left(e^{-\Delta_\lambda T} + \mu\left(e^{\Delta_\lambda T} - e^{-\Delta_\lambda T}\right)\right)} < 0.$$

When T is given by (10), we have

$$\frac{dT}{d\mu} = \frac{1}{\Delta_\lambda}\frac{c''(\mu)(k + c(\mu))}{(k + c(\mu) + (1-\mu)c'(\mu))(k + c(\mu) - c'(\mu)\mu)} > 0.$$

Finally, when T is determined from (11), we use

$$f(\mu, T) = \Delta_r\left(1 - e^{-\Delta_\lambda T}\right) + \Delta_r\mu\left(e^{\Delta_\lambda T} + e^{-\Delta_\lambda T} - 2\right) - \Delta_\lambda\left(1 - e^{-\Delta_r T}\right),$$

so that the optimal $T = T(\mu)$ solves $f(\mu, T(\mu)) = 0$. Then,

$$\left.\frac{dT}{d\mu}\right|_{T=T(\mu)} = -\frac{\partial f/\partial\mu|_{T=T(\mu)}}{\partial f/\partial T|_{T=T(\mu)}} < 0,$$

where we have used that $\partial f/\partial\mu = \Delta_r e^{-\Delta_\lambda T}\left(e^{\Delta_\lambda T} - 1\right)^2 > 0$ and $\partial f/\partial T|_{T=T(\mu)} > 0$, which follows from the arguments in the proof of Propositions 1 and 3.[31] This further implies that b_0, as defined by (7), must be decreasing in μ:

$$\frac{db_0}{d\mu} = c'(\mu)\frac{\lambda_H e^{\lambda_H T^*}}{\left(e^{\lambda_H T^*} - 1\right)^2}\left.\frac{dT}{d\mu}\right|_{T=T^*} - c''(\mu)\left(\mu + \frac{1}{e^{\lambda_H T^*} - 1}\right) < 0.$$

Q.E.D.

[31]In particular, Lemma A1 shows that, when $k > \overline{k}$ and (6) holds, then $f(T)$ has a unique interior (non-zero) solution T^* and is sloping upwards at T^*.

Proof of Proposition 2. We first complete the characterization. We denote by μ_1^*, μ_2^*, and μ_3^* the diligence level that the bank would optimally implement if regimes 1-3 applied and the compensation cost function $W(\mu)$ was determined accordingly. (Hence, for this auxiliary step it is not necessary to check whether the assumptions of the respective regime indeed hold for the chosen value of μ.) The respective contractual parameters are indexed accordingly. Define then, in complete analogy to the thresholds in (12) and (13),

$$\begin{aligned}\underline{k}^* &= c'(\mu_1^*)\left(\mu_1^* + \frac{1}{e^{\Delta_\lambda T_1^*} - 1}\right) - c(\mu_1^*), \\ \overline{k}^* &= c'(\mu_3^*)\left(\mu_3^* + \frac{1}{e^{\Delta_\lambda T_3^*} - 1}\right) - c(\mu_3^*).\end{aligned}$$

Finally, we make also the condition for when regime 3 applies more explicit:

$$\frac{\Delta_r}{\Delta_\lambda} < \gamma = 1 - 2\mu_3^*. \tag{27}$$

We now turn to the proof of the proposition.

The equilibrium wage cost function is given by:

$$W(\mu) = \begin{cases} c(\mu) + k + c'(\mu)\left(e^{\Delta_r T_3(\mu)} - 1\right)\left[\mu + \frac{1}{e^{\Delta_\lambda T_3(\mu)} - 1}\right] & \text{for} \quad \mu < \overline{k}^{-1}(\mu) \\ \left(1 + \frac{c'(\mu)}{k + c(\mu) - c'(\mu)\mu}\right)^{\frac{\Delta_r}{\Delta_\lambda}}\left[k + c(\mu)\right] & \text{for} \quad \overline{k}^{-1}(\mu) < \mu < \underline{k}^{-1}(\mu) \\ c'(\mu)\, e^{\Delta_r T_1(\mu)}\left[\mu + \frac{1}{e^{\Delta_\lambda T_1(\mu)} - 1}\right] & \text{for} \quad \mu > \underline{k}^{-1}(\mu) \end{cases}, \tag{28}$$

corresponding to regime 3 ($\mu < \overline{k}^{-1}(\mu)$), regime 2 ($\overline{k}^{-1}(\mu) < \mu < \underline{k}^{-1}(\mu)$) and regime 1 ($\mu > \underline{k}^{-1}(\mu)$) respectively. We first derive two auxiliary results. Consider the following maximization problems:

$$\mu_1^* = \arg\max\left\{\pi(\mu) - W_1(\mu)\right\}, \text{ s.t. } 0 \leq \mu < 1, \tag{29}$$

$$\mu_3^* = \arg\max\left\{\pi(\mu) - W_3(\mu)\right\}, \text{ s.t. } 0 \leq \mu \leq \tilde{\mu} := \frac{1}{2}\left(1 - \frac{\Delta_r}{\Delta_\lambda}\right), \tag{30}$$

where $W_i(\mu)$ refers to the wage cost function $W(\mu)$ in regime i (see 28). Note that both problems (29) and (30) are independent of k and well defined *regardless* of the equilibrium regime: The domains of these maximization problems are exogenous and, hence, unaffected by the endogenous equilibrium regimes. By assumption (strict quasiconcavity and $\lim_{\mu\to 1} c'(\mu)$ sufficiently high), the unique solution μ_1^* satisfies the first-order condition

$$\pi'(\mu_1^*) = W_1'(\mu_1^*).$$

Instead, μ_3^* is either given by first-order conditions or its maximum (corner) value $\tilde{\mu} = \frac{1}{2}\left(1 - \frac{\Delta_r}{\Delta_\lambda}\right)$.[32] Note that $T_3(\mu)$ is positive for any $\mu < \tilde{\mu}$ and satisfies $\lim_{\mu\to\tilde{\mu}} T_3(\mu) = 0$. Since $\lim_{\mu\to\tilde{\mu}} W_3'(\mu) = c'(\tilde{\mu}) + c''(\tilde{\mu})\frac{\Delta_r}{\Delta_\lambda}$, a corner solution obtains if $\pi'(\tilde{\mu}) \geq c'(\tilde{\mu}) + c''(\tilde{\mu})\frac{\Delta_r}{\Delta_\lambda}$. In this case, the auxiliary problem implies that regime 3 can never obtain in equilibrium (for any level of k).

In the following, we consider the relevant case when $\pi'(\tilde{\mu}) < c'(\tilde{\mu}) + c''(\tilde{\mu})\frac{\Delta_r}{\Delta_\lambda}$. Denote then the interior solution μ_3 as μ_3^* which is characterized by the first-order condition

$$\pi'(\mu_3^*) = W_3'(\mu_3^*).$$

Lemma A2. $\mu_1^* < \mu_3^*$.

Proof. For any $0 \leq \mu < \tilde{\mu}$, $\mu_3^* > \mu_1^*$ holds if $W_1'(\mu) > W_3'(\mu)$. Using the envelope theorem we obtain:

$$\begin{aligned} W_3'(\mu) &= c'(\mu) \\ &\quad + c''(\mu)\left(e^{\Delta_r T_3} - 1\right)\left[\mu + \frac{1}{e^{\Delta_\lambda T_3} - 1}\right] + c'(\mu)\left(e^{\Delta_r T_3} - 1\right) \end{aligned}$$

and

$$\begin{aligned} W_1'(\mu) &= c'(\mu) + c''(\mu)\left[\mu + \frac{1}{e^{\Delta_\lambda T_1} - 1}\right] \\ &\quad + c''(\mu)\left(e^{\Delta_r T_1} - 1\right)\left[\mu + \frac{1}{e^{\Delta_\lambda T_1} - 1}\right] + c'(\mu)\left(e^{\Delta_r T_1} - 1\right). \end{aligned}$$

In order to show that $W_1'(\mu) > W_3'(\mu)$, it is clearly sufficient to compare the second lines in the respective expressions. The assertion then follows from the following two observations. First, from $T_3 < T_1$ we have $e^{\Delta_r T_3} < e^{\Delta_r T_1}$. Second, note that by the definition of T_3 the expression $\left(e^{\Delta_r T} - 1\right)\left[\mu + \frac{1}{e^{\Delta_\lambda T} - 1}\right]$ is minimized at T_3. **Q.E.D.**

Lemma A3. $\underline{k}^* < \overline{k}^*$.

Proof. If $\mu_3^* = \tilde{\mu}$, we set $\overline{k}^* = \infty$ and the relationship trivially holds. Now consider the case when $\mu_3^* < \tilde{\mu}$. Recall that T_1 and T_3 both decrease in μ. As for given μ we have $T_3 < T_1$, we thus have $T_3(\mu_3^*) < T_1(\mu_1^*)$, as determined at the respective optimal choices for μ. Now, consider the function $\tilde{k}(\mu, T) = c'(\mu)\left(\mu + \frac{1}{e^{\Delta_\lambda T} - 1}\right) - c(\mu)$. Since $\tilde{k}$ is increasing in μ and decreasing in T, it must be true that $\overline{k}^* = \tilde{k}(\mu_3^*, T_3(\mu_3^*)) > \underline{k}^* = \tilde{k}(\mu_1^*, T_1(\mu_1^*))$. **Q.E.D.**

[32] Thus, we directly get that γ as defined in (27) is strictly positive.

Take now the first assertion in Proposition 2. If we solve the relaxed program (ignoring the acquisition constraint) and this solution automatically satisfies the acquisition constraint, then the relaxed program also solves the full program. Put differently, if $k < \underline{k}^*$, then regime 1 obtains in equilibrium and $\mu^* = \mu_1^*$. Now, we consider the case where $k > \underline{k}^*$ and the acquisition constraint binds, i.e., regime 1 does not obtain and either regime 2 or 3 occur. If either $\frac{\Delta_r}{\Delta_\lambda} > 1$ or $\pi'(\tilde{\mu}) \geq c'(\tilde{\mu}) + c''(\tilde{\mu}) \frac{\Delta_r}{\Delta_\lambda}$, then regime 3 is not feasible and regime 2 obtains for any $k > \underline{k}^*$. If $\pi'(\tilde{\mu}) < c'(\tilde{\mu}) + c''(\tilde{\mu}) \frac{\Delta_r}{\Delta_\lambda}$, then regime 3 does not violate the constraint $b_0 \geq 0$ provided that $k \geq \overline{k}^*$. Whenever regime 3 is feasible, it is preferable to the constrained regime 2. Since $\overline{k}^* > \underline{k}^*$, this implies that regime 3 obtains if $k \geq \overline{k}^*$. **Q.E.D.**

Proof of Lemma 2. Denote the time of the first strictly positive payment by $T_0 \geq \tau$ and substitute out the associated payment $b_0 > 0$ from the two incentive constraints (2) and (3) to get the requirement

$$\begin{aligned}\sum\nolimits_{i\geq 1} b_i e^{-(r_A+\lambda_H)T_i}\left[e^{\Delta_\lambda T_i} - e^{\Delta_\lambda T_0}\right] &\leq c'(\mu) - \left(e^{\Delta_\lambda T_0} - 1\right)(k + c(\mu) - \mu c'(\mu)) \\ &\leq c'(\mu) - \left(e^{\Delta_\lambda \tau} - 1\right)(k + c(\mu) - \mu c'(\mu)),\end{aligned}$$

where the second inequality follows from $T_0 \geq \tau$ and $k > \mu c'(\mu) - c(\mu)$. Now note that the left-hand side is non-negative as T_0 was defined as the time of the first strictly positive payment, while the right-hand-side becomes zero for $\tau = T_2$ as defined in (10) and negative for $\tau > T_2$. Hence, the two constraints (2) and (3) can only be satisfied using non-negative payments at times $T_i \geq \tau$ if $\tau \leq T_2$. Finally, the comparative statics result in k follows from inspection of (10), and the positive dependence on μ from Corollary 2**.**
Q.E.D.

Proof of Proposition 4. With $\tau > 0$, the optimal level of diligence effort for regime 1 solves the following program:

$$\begin{aligned}\max_{\mu} \{\pi(\mu) - W_1(\mu, t(\mu))\} \\ s.t.\, t(\mu) \geq \tau \\ \mu \geq \underline{\mu}(\tau),\end{aligned} \tag{31}$$

where the lower bound on μ is uniquely determined from

$$k + c(\mu) - \mu c'(\mu) = \frac{c'(\mu)}{e^{\Delta_\lambda \tau} - 1}, \tag{32}$$

if this admits a positive solution, while we set $\underline{\mu}(\tau) = 0$ else.[33] From Propositions 1 and 3, we further have that $t(\mu) = T_1(\mu)$ as determined from (9) for $T_1(\mu) \geq \tau$, while else it holds that $t(\mu) = \tau$. Now, from

$$W_1(\mu, t) = c'(\mu) \, e^{\Delta_r t} \left[\mu + \frac{1}{e^{\Delta_\lambda t} - 1} \right],$$

it holds that

$$\frac{\partial^2 W_1}{\partial t \partial \mu} = \frac{c''(\mu)}{c'(\mu)} \frac{\partial W_1}{\partial t} + c'(\mu) \Delta_r e^{\Delta_r t} > 0 \text{ for } t > T_1,$$

where we have used that $\partial W_1 / \partial t > 0$ for $t > T_1$ (cf. the proof of Proposition 1). Hence, the optimally implemented level of diligence, $\mu_1^*(\tau)$, must be decreasing in τ for $\tau \geq T_1(\mu_1^*(0))$ as long as $\mu_1^*(\tau) > \underline{\mu}(\tau)$.[34] Together with

$$\frac{\partial \underline{\mu}(\tau)}{\partial \tau} = \frac{c'(\mu) \, \frac{\Delta_\lambda e^{\Delta_\lambda \tau}}{\left(e^{\Delta_\lambda \tau} - 1\right)^2}}{\mu c''(\mu) + \frac{c''(\mu)}{e^{\Delta_\lambda \tau} - 1}} > 0,$$

this implies that there exists $\widetilde{\tau}$, such that $\mu_1^*(\tau) = \underline{\mu}(\tau)$ for all $\tau \geq \widetilde{\tau}$. Hence, $\mu_1^*(\tau)$ is increasing in τ for all $\tau \geq \widetilde{\tau}$ up to $\overline{\mu}$, which, thus, is the maximal value of diligence that can be achieved with any $\tau > T_1(\mu_1^*(0))$. **Q.E.D.**

Proof of Proposition 5. We will first prove the claim for the case where, initially, i.e., for $\tau = 0$, regime 2 applies. So assume that either $k \geq \underline{k}^*$ and (6) is violated, or, that $\underline{k}^* \leq k \leq \overline{k}^*$ and (6) holds. Then, the optimal level of diligence $\mu_2^*(\tau)$ is still determined from program (31), where now we have $t(\mu) = T_2(\mu)$ as given by (10) as long as $T_2(\mu) \leq \tau$ and $t(\mu) = \tau$ else. Thus, the regulatory constraint becomes binding for $\tau > T_2(\mu_2^*(0))$, in which case it must hold that $\mu_2^*(\tau) = \underline{\mu}(\tau)$, which is increasing in τ. To see this assume to the contrary that $\mu_2^*(\tau) > \underline{\mu}(\tau)$, for some $\tau > T_2(\mu_2^*(0))$. This implies, however, that the acquisition constraint is slack such that we are back to regime 1. Then the arguments in the proof of Proposition 4 above, together with the observation that, by definition, $\mu_2^*(0) = \underline{\mu}(T_2(\mu_2^*(0)))$, imply that $\mu_2^*(\tau) = \underline{\mu}(\tau)$, contradiction.

Next, consider the case where $k > \overline{k}^*$ and (6) holds, such that regime 3 applies for $\tau = 0$. Here, we distinguish two cases, depending on whether there are two payments

[33] Here we again use that $c(\cdot)$ is sufficiently convex such that $k + c(1) - c'(1) \leq 0$, and, hence, $\underline{\mu}(\tau)$ is for all τ determined by the solution to (32) as long as this is non-negative. Existence of a positive solution to (32) is guaranteed for $\tau > \tau'$, where $\tau' = \frac{1}{\Delta_\lambda} \ln \left(1 + \frac{c'(0)}{k}\right)$ stays bounded, as long as $k > 0$.

[34] From the assumption that $k < \underline{k}^*$ this region is non-empty.

also with regulation or only one. So assume, first, that also with regulation, there are two payments at the optimally implemented level of diligence. Then, from Proposition 3, the constraint $\mu \geq \underline{\mu}(\tau)$ does not bind. Also, we have $b_\tau > 0$ and $b_T > 0$. It is then useful to define the following function:[35]

$$\begin{aligned} w(\mu, T, \tau) &:= e^{\Delta_r T} c'(\mu) \\ &+ \left(e^{\Delta_r T} - e^{\Delta_r \tau}\right) \left(\frac{\left(1 + \mu\left(e^{\Delta_\lambda T} - 1\right)\right)\left(1 + \mu\left(e^{\Delta_\lambda \tau} - 1\right)\right)}{e^{\Delta_\lambda T} - e^{\Delta_\lambda \tau}} c''(\mu) - \frac{e^{\Delta_\lambda \tau} - 1}{e^{(r_A + \lambda_H)\tau}} b_\tau \right), \end{aligned}$$

which satisfies

$$w\left(\mu_3^*, T_3^*(\tau), \tau\right) = dW_3/d\mu|_{T=T(\mu)}.$$

Thus, the function w just represents the marginal wage cost if evaluated at the optimum T_3^*. The equilibrium choice μ_3^* in regime 3 then satisfies

$$\pi'\left(\mu_3^*\right) = w\left(\mu_3^*, T_3^*, \tau\right)$$

Since $\frac{dw}{d\tau} < 0$ implies the claim that $\frac{d\mu_3^*}{d\tau} > 0$, we need to show that:

$$\frac{dw}{d\tau} = \frac{\partial w}{\partial T}\frac{\partial T}{\partial \tau} + \frac{\partial w}{\partial \tau} < 0.$$

Further, as, from the proof of Proposition 3, $\frac{\partial T}{\partial \tau} < 0$, it suffices to show that $\frac{\partial w}{\partial \tau} < 0$ and $\frac{\partial w}{\partial T} > 0$. Let us, first, consider $\frac{\partial w}{\partial T}$, which is given by

$$\frac{\partial w}{\partial T} = \frac{e^{-(r_A + \lambda_H)T} b_T}{e^{\Delta_\lambda T} - e^{\Delta_\lambda \tau}} \Xi_1 + \frac{\left(1 + \mu\left(e^{\Delta_\lambda \tau} - 1\right)\right) c''(\mu)}{\left(e^{\Delta_\lambda T} - e^{\Delta_\lambda \tau}\right)^2} \Xi_2,$$

where the terms Ξ_1 and Ξ_2 are defined as

$$\begin{aligned} \Xi_1 &= \Delta_r e^{\Delta_r T}\left(e^{\Delta_\lambda T} - e^{\Delta_\lambda \tau}\right)\left(e^{\Delta_\lambda T} - 1\right) - \Delta_\lambda e^{\Delta_\lambda T}\left(e^{\Delta_r T} - e^{\Delta_r \tau}\right)\left(e^{\Delta_\lambda \tau} - 1\right), \\ \Xi_2 &= \Delta_r e^{\Delta_r T}\left(e^{\Delta_\lambda T} - e^{\Delta_\lambda \tau}\right)\left(1 + \mu\left(e^{\Delta_\lambda T} - 1\right)\right) - \Delta_\lambda e^{\Delta_\lambda T}\left(e^{\Delta_r T} - e^{\Delta_r \tau}\right)\left(1 + \mu\left(e^{\Delta_\lambda \tau} - 1\right)\right). \end{aligned}$$

Using the optimality condition for T (cf. 16) these expressions can be simplified to obtain

$$\begin{aligned} \Xi_1 &= \frac{\Delta_r e^{\Delta_r T}\left(e^{\Delta_\lambda T} - e^{\Delta_\lambda \tau}\right)^2}{1 + \mu\left(e^{\Delta_\lambda \tau} - 1\right)} > 0, \\ \Xi_2 &= 0. \end{aligned}$$

[35] Here, b_τ formally represents a function $b_\tau(\mu, T, \tau)$.

Since $b_T > 0$ and $\Xi_1 > 0$, it follows that $\frac{\partial w}{\partial T} > 0$. Next, consider $\frac{\partial w}{\partial \tau}$ as given by

$$\frac{\partial w}{\partial \tau} = \frac{e^{-(r_A+\lambda_H)\tau} b_\tau}{e^{\Delta_\lambda T} - e^{\Delta_\lambda \tau}} \Xi_3 + \frac{\left(1 + \mu \left(e^{\Delta_\lambda T} - 1\right)\right) c''(\mu)}{\left(e^{\Delta_\lambda T} - e^{\Delta_\lambda \tau}\right)^2} \Xi_4,$$

where the terms Ξ_3 and Ξ_4 are defined as

$$\begin{aligned}
\Xi_3 &= \Delta_r e^{\Delta_r \tau} \left(e^{\Delta_\lambda T} - e^{\Delta_\lambda \tau}\right) \left(e^{\Delta_\lambda \tau} - 1\right) - \Delta_\lambda e^{\Delta_\lambda \tau} \left(e^{\Delta_r T} - e^{\Delta_r \tau}\right) \left(e^{\Delta_\lambda T} - 1\right), \\
\Xi_4 &= \Delta_r e^{\Delta_r \tau} \left(e^{\Delta_\lambda T} - e^{\Delta_\lambda \tau}\right) \left(1 + \mu \left(e^{\Delta_\lambda \tau} - 1\right)\right) \\
&\quad + \Delta_\lambda e^{\Delta_\lambda \tau} \left(e^{\Delta_r T} - e^{\Delta_r \tau}\right) \left(1 + \mu \left(e^{\Delta_\lambda T} - 1\right)\right).
\end{aligned}$$

Using again the optimality condition for T (cf. 16) these expressions can be rewritten to obtain

$$\begin{aligned}
\Xi_3 &= -\Delta_r e^{\Delta_r \tau} \left(e^{\Delta_\lambda T} - e^{\Delta_\lambda \tau}\right) \left(e^{\Delta_\lambda \tau} - 1\right) \left[\frac{e^{\Delta_r T}}{e^{\Delta_r \tau}} \frac{e^{\Delta_\lambda \tau} - e^{-\Delta_\lambda (T-\tau)}}{e^{\Delta_\lambda \tau} - 1} \frac{1 + \mu \left(e^{\Delta_\lambda T} - 1\right)}{1 + \mu \left(e^{\Delta_\lambda \tau} - 1\right)} - 1\right], \\
\Xi_4 &= -\Delta_r e^{\Delta_r \tau} \left(e^{\Delta_\lambda T} - e^{\Delta_\lambda \tau}\right) \left(1 + \mu \left(e^{\Delta_\lambda \tau} - 1\right)\right) \left[1 - \frac{\Delta_\lambda^2}{\Delta_r^2} \frac{e^{\Delta_\lambda (T-\tau)}}{e^{\Delta_r (T-\tau)}} \left(\frac{e^{\Delta_r (T-\tau)} - 1}{e^{\Delta_\lambda (T-\tau)} - 1}\right)^2\right],
\end{aligned}$$

where we find that $\Xi_3 < 0$ as the numerator of each factor of the term in square brackets is greater than the respective denominator (by $T > \tau$) such that their product must be greater than 1. To see that $\Xi_4 < 0$, one has to prove that $\varphi(x) = 1 - \frac{\Delta_\lambda^2}{\Delta_r^2} \frac{e^{\Delta_\lambda x}}{e^{\Delta_r x}} \left(\frac{e^{\Delta_r x} - 1}{e^{\Delta_\lambda x} - 1}\right)^2$ is non-negative for $x > 0$. When $\Delta_\lambda > \Delta_r$, the function $\varphi(x)$ satisfies: $\inf \varphi(x) = \lim_{x \to 0} \varphi(x) = 0$. As a result, $\varphi(x) > 0$ for $x > 0$. Taken together, the conditions $\Xi_3 < 0$, $\Xi_4 < 0$, and $b_\tau > 0$ imply that $\frac{\partial w}{\partial \tau} < 0$. Hence, we have shown, for the case with two payments, that indeed $\mu_3^*(\tau)$ is increasing in τ.

Finally, assume that with regulation there is only a single payment at the optimally implemented level of diligence. Then the regulatory constraint on μ binds, i.e., $\mu^*(\tau) = \underline{\mu}(\tau)$, which is increasing in τ. **Q.E.D.**

Proof of Proposition 6. If $k \geq \underline{k}^*$, it holds from Proposition 5 that $\mu^*(\tau)$ is strictly increasing in τ if the regulatory constraint is binding and, thus, maximized at the highest possible value $\tau = \overline{\tau}$. If $k < \underline{k}^*$, we know from Proposition 4 that $\mu_1^*(\tau)$ is, for $\tau < \widetilde{\tau}$, decreasing and, for $\tau > \widetilde{\tau}$, increasing in τ. Hence, it must be maximized either for $\tau \in [0, T_1(\mu_1^*(0))]$, i.e., when the regulatory constraint does not bind, or at $\tau = \overline{\tau}$. Consider first the case where the bank's zero profit constraint can be ignored, i.e., $\overline{\tau} \to \infty$. Then diligence is maximized for $\tau \to \infty$ if and only if $\mu_1^*(0) < \lim_{\tau \to \infty} \mu_1^*(\tau) = \overline{\mu}$. Now, from (14) it holds that $\overline{\mu}$ is strictly increasing in k and approaches zero for $k \to 0$, while $\mu_1^*(0)$ is, from the slack acquisition constraint, independent of k. Hence, there exists a unique value $\widetilde{k} = c'(\mu_1^*(0))\mu_1^*(0) - c(\mu_1^*(0))$, such that $\mu_1^*(0) \geq \overline{\mu}$ for $k \leq \widetilde{k}$, and $\mu_1^*(0) < \overline{\mu}$ for $k > \widetilde{k}$, with $\widetilde{k} < \underline{k}^*$.

When the bank's zero-profit constraint has to be satisfied, the only difference is that, for $k < \underline{k}^*$, the relevant comparison now is between $\mu_1^*(0)$ and $\mu_1^*(\overline{\tau})$. Trivially, as, for $\overline{\tau} < \infty$, $\mu_1^*(0) \geq \overline{\mu}$ implies $\mu_1^*(0) > \mu_1^*(\overline{\tau})$, diligence for $k \leq \widetilde{k}$ is still maximized when no binding deferral regulation is imposed. It remains to deal with the case $\widetilde{k} < k < \underline{k}^*$.

If μ is maximized at $\tau = \overline{\tau}$, given $\widetilde{k} < k < \underline{k}^*$, it must hold that $\mu_1^*(\overline{\tau}) = \underline{\mu}(\overline{\tau}(k), k)$, making explicit the dependence on k. Note now that

$$\begin{aligned} \frac{d}{dk}\underline{\mu}(\overline{\tau}(k), k) &= \frac{\partial \underline{\mu}(\overline{\tau}, k)}{\partial k} + \frac{\partial \underline{\mu}(\overline{\tau}, k)}{\partial \tau}\overline{\tau}'(k) \\ &= \frac{1}{c''(\underline{\mu})\left(\underline{\mu} + \frac{1}{e^{\Delta_\lambda \overline{\tau}}-1}\right)}\left[1 + c'(\underline{\mu})\frac{\Delta_\lambda e^{\Delta_\lambda \overline{\tau}}}{(e^{\Delta_\lambda \overline{\tau}} - 1)^2}\overline{\tau}'(k)\right], \end{aligned} \tag{33}$$

where we have used that, from (32),

$$\begin{aligned} \frac{\partial \underline{\mu}(\tau, k)}{\partial \tau} &= \frac{c'\left(\underline{\mu}\right)\frac{\Delta_\lambda e^{\Delta_\lambda \tau}}{\left(e^{\Delta_\lambda \tau}-1\right)^2}}{c''(\underline{\mu})\left(\underline{\mu} + \frac{1}{e^{\Delta_\lambda \tau}-1}\right)}, \\ \frac{\partial \underline{\mu}(\tau, k)}{\partial k} &= \frac{1}{c''(\underline{\mu})\left(\underline{\mu} + \frac{1}{e^{\Delta_\lambda \tau}-1}\right)}. \end{aligned}$$

Next, the zero profit condition

$$\pi(\underline{\mu}(\overline{\tau}, k)) - e^{\Delta_r \overline{\tau}}c'\left(\underline{\mu}(\overline{\tau}, k)\left[\underline{\mu}(\overline{\tau}, k) + \frac{1}{e^{\Delta_\lambda \overline{\tau}} - 1}\right] = 0$$

implies

$$\overline{\tau}'(k) = -\frac{\left(\pi'(\underline{\mu}) - e^{\Delta_r \overline{\tau}}c'(\underline{\mu})\right)\frac{1}{c''(\underline{\mu})\left(\underline{\mu}+\frac{1}{e^{\Delta_\lambda \overline{\tau}}-1}\right)} - e^{\Delta_r \overline{\tau}}}{\left(\pi'(\underline{\mu}) - e^{\Delta_r \overline{\tau}}c'(\underline{\mu})\right)\frac{c'(\underline{\mu})\frac{\Delta_\lambda e^{\Delta_\lambda \overline{\tau}}}{\left(e^{\Delta_\lambda \overline{\tau}}-1\right)^2}}{c''(\underline{\mu})\left(\underline{\mu}+\frac{1}{e^{\Delta_\lambda \overline{\tau}}-1}\right)} - \Delta_r e^{\Delta_r \overline{\tau}}c'(\underline{\mu})\left[\underline{\mu} + \frac{1}{e^{\Delta_\lambda \overline{\tau}}-1}\right]},$$

where we note that both the numerator as well as the denominator are negative. For the numerator we have

$$\pi'(\underline{\mu}) - e^{\Delta_r \overline{\tau}}\left(c'(\underline{\mu}) + c''(\underline{\mu})\left(\underline{\mu} + \frac{1}{e^{\Delta_\lambda \overline{\tau}} - 1}\right)\right) < 0$$

which follows from the observation that $\underline{\mu}$ is the lowest value of diligence that can be "implemented", and, is, in particular, not determined by a first-order condition. For the denominator we have

$$\left(\pi'(\underline{\mu}) - e^{\Delta_r \overline{\tau}} c''(\underline{\mu}) \left[\underline{\mu} + \frac{1}{e^{\Delta_\lambda \overline{\tau}} - 1}\right] - e^{\Delta_r \overline{\tau}} c'(\underline{\mu})\right) \frac{\partial \underline{\mu}(\overline{\tau}, k)}{\partial \overline{\tau}}$$
$$- e^{\Delta_r \overline{\tau}} c'(\underline{\mu}) \left(\Delta_r \left[\underline{\mu} + \frac{1}{e^{\Delta_\lambda \overline{\tau}} - 1}\right] - \frac{\Delta_\lambda e^{\Delta_\lambda \overline{\tau}}}{\left(e^{\Delta_\lambda \overline{\tau}} - 1\right)^2}\right)$$

where the first term is again negative. As for the second term we obtain

$$\begin{aligned}
& e^{\Delta_r \overline{\tau}} c'(\underline{\mu}) \left(\Delta_r \left[\underline{\mu} + \frac{1}{e^{\Delta_\lambda \overline{\tau}} - 1}\right] - \frac{\Delta_\lambda e^{\Delta_\lambda \overline{\tau}}}{\left(e^{\Delta_\lambda \overline{\tau}} - 1\right)^2}\right) \\
&= e^{\Delta_r \overline{\tau}} \frac{1}{\left(e^{\Delta_\lambda \overline{\tau}} - 1\right)^2} c'(\underline{\mu}) \left(\Delta_r \left[\underline{\mu} \left(e^{\Delta_\lambda \overline{\tau}} - 1\right)^2 + \left(e^{\Delta_\lambda \overline{\tau}} - 1\right)\right] - \Delta_\lambda e^{\Delta_\lambda \overline{\tau}}\right) \\
&= e^{\Delta_r \overline{\tau}} e^{\Delta_\lambda \overline{\tau}} \frac{1}{\left(e^{\Delta_\lambda \overline{\tau}} - 1\right)^2} c'(\underline{\mu}) \underbrace{\left(\Delta_r \left(1 - e^{-\Delta_\lambda \overline{\tau}}\right) + \Delta_r \underline{\mu} \left(e^{\Delta_\lambda \overline{\tau}} - 2 + e^{-\Delta_\lambda \overline{\tau}}\right) - \Delta_\lambda\right)}_{>0} \\
&> 0,
\end{aligned}$$

where the inequality follows from Lemma A1 (in particular, for $T > T_1$ the left-hand-side of (25) is always strictly positive). Thus, $\frac{d}{dk} \underline{\mu}(\overline{\tau}(k), k)$ as given in (33) is positive if and only if

$$\begin{aligned}
& 1 + c'(\underline{\mu}) \frac{\Delta_\lambda e^{\Delta_\lambda \overline{\tau}}}{\left(e^{\Delta_\lambda \overline{\tau}} - 1\right)^2} \overline{\tau}'(k) > 0 \\
\iff & 1 - \frac{\left(\pi'(\underline{\mu}) - e^{\Delta_r \overline{\tau}} c'(\underline{\mu})\right) \frac{1}{c''(\underline{\mu})\left(\underline{\mu} + \frac{1}{e^{\Delta_\lambda \overline{\tau}} - 1}\right)} - e^{\Delta_r \overline{\tau}}}{\left(\pi'(\underline{\mu}) - e^{\Delta_r \overline{\tau}} c'(\underline{\mu})\right) \frac{1}{c''(\underline{\mu})\left(\underline{\mu} + \frac{1}{e^{\Delta_\lambda \overline{\tau}} - 1}\right)} - \frac{\Delta_r e^{\Delta_r \overline{\tau}}}{\Delta_\lambda e^{\Delta_\lambda \overline{\tau}}} \left[\underline{\mu} \left(e^{\Delta_\lambda \overline{\tau}} - 1\right)^2 + \left(e^{\Delta_\lambda \overline{\tau}} - 1\right)\right]} > 0 \\
\iff & \Delta_r \left(1 - e^{-\Delta_\lambda \overline{\tau}}\right) + \Delta_r \underline{\mu} \left(e^{\Delta_\lambda \overline{\tau}} + e^{-\Delta_\lambda \overline{\tau}} - 2\right) - \Delta_\lambda > 0,
\end{aligned}$$

which again holds from Lemma A1 (in particular for $T > T_1$ the left-hand-side of (25) is strictly positive). Hence, there exists a unique cutoff $\widehat{k} < \underline{k}^*$ such that $\underline{\mu}(\overline{\tau}(k), k) < \mu^*(0)$ for $k < \widehat{k}$ and $\underline{\mu}(\overline{\tau}(k), k) > \mu^*(0)$ for $k > \widehat{k}$. **Q.E.D.**

Proof of Proposition 7. The proof relies repeatedly on the following auxiliary result.

Lemma A4. The effect of binding regulation via mandatory deferral on $\Pi + V_A$ is always strictly negative, i.e., $\frac{d}{d\tau} \left(\Pi + V_A\right) < 0$ as long as $\tau < \overline{\tau}$.

Proof. As the bank maximizes Π for any given τ, clearly, $d\Pi/d\tau$ is strictly negative for any $\tau < \overline{\tau}$ where deferral regulation is binding. As for the effect on V_A take, first, the case where the acquisition constraint is binding. Then, $V_A = 0$ and the result follows. It remains to consider the case where the acquisition constraint is slack. In this case we know from Proposition 4 that $\mu^*(\tau)$ is strictly decreasing in τ, if deferral regulation is binding. The result then follows as V_A is strictly increasing in $\mu^*(\tau)$ and strictly decreasing in τ:

$$\frac{dV_A}{d\tau} = -c'\left(\mu^*(\tau)\right)\frac{\Delta_\lambda e^{\Delta_\lambda \tau}}{(e^{\Delta_\lambda \tau}-1)} + \left(c''\left(\mu^*(\tau)\right)\left[\mu^*(\tau) + \frac{1}{e^{\Delta_\lambda \tau}-1}\right]\right)\frac{\partial \mu^*(\tau)}{\partial \tau} < 0.$$

Q.E.D.

Next, we will show that $k > \widehat{k}$ and X sufficiently high are both necessary conditions for the optimality of a binding deferral regulation. Assume to the contrary that welfare Ω_h is maximized for some (binding) τ while $k \leq \widehat{k}$. In this case we know from Proposition 6 that $\mu^*(\tau) < \mu^*(0)$, which from (18) leads to a higher expected externality. Together with the from Lemma A4 negative effect on $\Pi + V_A$, it then must hold from (19) that $\Omega_h(\tau) < \Omega_h(0)$, contradiction. Thus, we have shown that $k > \widehat{k}$ is necessary for binding deferral regulation to maximize welfare. Finally, the necessity of a sufficiently high externality X follows trivially from Lemma A4 together with the fact that Ω_h is continuous and increasing in X and satisfies, for $X = 0$, $\Omega_h = \Pi + V_A$.

As for the comparative analysis in X, differentiating (19) with respect to τ and X results in

$$\frac{d^2\Omega_h}{d\tau dX} = -\frac{d\mu^*(\tau)}{d\tau}\frac{d^2\omega}{d\mu dX} = \left(\frac{\lambda_H}{r+\lambda_H} - \frac{\lambda_L}{r+\lambda_L}\right)\frac{d\mu^*(\tau)}{d\tau},$$

where the last term is in the relevant range strictly positive. Thus, it follows from standard monotone comparative statics results that the optimal deferral time is strictly increasing in X. **Q.E.D.**

Appendix B: Compensation Design with Continuous Acquisition Effort

In this Appendix we consider optimal compensation design without regulation in a setting where also acquisition effort is a continuous variable. Notably, most of the characterization results obtained in the main text for the case of binary acquisition effort, can be applied one-to-one also to the case where a is a continuous choice. Then, through exerting unobservable acquisition effort $a \in [0,1]$ at private disutility $k(a)$, the agent generates a business opportunity with probability a. We further stipulate that the respective cost function $k(a)$ is twice continuously differentiable with $k''(a) > 0$, $k'(0) = 0$ and $k'(a)$ sufficiently large as $a \to 1$. The remaining model specifications are as in the main text. Hence, the agent's discounted expected payoff for a given choice of acquisition effort a and diligence effort μ is still given by (1). Now, to ensure that the given choice of a and μ is indeed optimal for the agent, the respective first-order conditions have to be satisfied: For μ we, thus, require (2) to hold, which we repeat here for convenience,

$$\sum\nolimits_i b_i e^{-r_A T_i} \left(e^{-\lambda_L T_i} - e^{-\lambda_H T_i}\right) = c'(\mu), \tag{34}$$

while for a the respective first-order condition is given by

$$\sum\nolimits_i b_i e^{-r_A T_i} \left[\mu e^{-\lambda_L T_i} + (1-\mu) e^{-\lambda_H T_i}\right] - c(\mu) = k'(a), \tag{35}$$

which replaces the inequality constraint for (binary) acquisition effort in (3). Total expected costs of compensation are then given by

$$W(a,\mu) = a \sum\nolimits_i b_i e^{-r_P T_i} \left[\mu e^{-\lambda_L T_i} + (1-\mu) e^{-\lambda_H T_i}\right], \tag{36}$$

and the bank's compensation design problem is to choose $b_i \geq 0$ and $T_i \geq 0$ to minimize (36) subject to the incentive constraints in (34) and (35). Comparing this program to the one with binary acquisition effort analyzed in the main text, the only difference is that the acquisition effort constraint is now given by the first-order condition in (35) and, thus, using the properties of $k(a)$, always binding. This also implies that certain combinations of a and μ are no longer implementable.

Lemma 4 *The combination of (a,μ) is implementable if and only if*

$$k'(a) + c(\mu) - c'(\mu)\mu \geq 0. \tag{37}$$

Proof of Lemma 4. Adding (34) and (35) we obtain the requirement

$$\sum_i b_i e^{-(r_A+\lambda_H)T_i} = k'(a) + c(\mu) - c'(\mu)\mu.$$

Since the left-hand side is non-negative from $b_i \geq 0$, any (a, μ) combination satisfying $k'(a)+c(\mu)-c'(\mu)\mu < 0$ is not implementable. If, however, (37) is satisfied, there clearly exists a choice of $b_i \geq 0$ and $T_i \geq 0$ such that both (34) and (35) are satisfied (cf. the optimal compensation contract characterized below). **Q.E.D.**

Intuitively, the implementability constraint in (37) arises from the complementarity of the agent's tasks of exerting acquisition and, subsequently, diligence effort. In particular, when the agent is induced to exert higher diligence, the resulting higher rent also creates incentives for acquisition. For any given level of μ, the minimal such diligence rent is given by $c'(\mu)\mu - c(\mu)$, which from (37) provides a lower bound on the level of acquisition effort a that can be implemented.[36]

The characterization of the optimal compensation contract provided in Proposition 1 and Lemma 1 now simply extends also to the case with continuos acquisition effort, once k is replaced in all expressions by the marginal costs $k'(a)$ evaluated at the level of acquisition effort a the principal wants to implement.[37] The only difference is, that, with binary a, the acquisition effort constraint was slack for low values $k < \underline{k}$ (regime 1). As acquisition effort a now has to satisfy a first-order condition (cf. (35)) this case can no longer occur. Indeed, given μ, it is impossible to implement particularly low values of a, satisfying $k'(a) < c'(\mu)\mu - c(\mu) < \underline{k}$. We, hence, have the following characterization result:

Proposition 8 *To implement a given level of acquisition effort a and diligence effort μ satisfying (37), the bank chooses a single, uniquely determined long-term bonus b_T, which satisfies (5) and a unique timing T. If (6) is satisfied and $k'(a) > \overline{k}$ as defined in (13), an additional up-front bonus $b_0 > 0$ given by (7) is paid. In this case T is determined from (11). In all other cases, i.e., when either (6) does not hold or $k'(a)$ is still sufficiently low with $c'(\mu)\mu - c(\mu) \leq k'(a) \leq \overline{k}$, there is no up-front bonus and T is given by (10).*

[36] The minimal diligence rent of $c'(\mu)\mu - c(\mu)$ arises from the fact that even if it was immediately observed whether λ_L or λ_H was realized, this is still just an imperfect signal about the agent's diligence effort μ (cf. also the discussion following Proposition 1 in the main text).

[37] Analogously, the solution to the compensation design problem under regulation, requiring $T_i \geq \tau$, analyzed in Section 4.1, also directly extends to the case with continuous acquisition effort.