DISCUSSION PAPER SERIES

No. 9877

REGULATING DEFERRED INCENTIVE PAY

Florian Hoffmann, Roman Inderst and Marcus Opp

FINANCIAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP9877.php**

ISSN 0265-8003

REGULATING DEFERRED INCENTIVE PAY

Florian Hoffmann, Johann Wolfgang Goethe-Universität Frankfurt
Roman Inderst, Johann Wolfgang Goethe-Universität Frankfurt, Imperial College London, and CEPR
Marcus Opp, University of California, Berkeley

Discussion Paper No. 9877
March 2014

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 9877

March 2014

ABSTRACT

Regulating Deferred Incentive Pay*

Our paper examines the effect of recent regulatory proposals mandating the deferral of bonus payments and claw-back clauses for compensation contracts in the financial sector. We study a multi-task setting in which a bank employee, the agent, privately chooses (deal or customer) acquisition effort and diligence, which stochastically reduces the occurrence of negative events over time (such as loan defaults or customer cancellations). The key ingredient of the compensation contract is the endogenous timing of a long-term bonus that trades off the cost and benefit of delay resulting from agent impatience and the informational gain, respectively. Our main finding is that government interference with this privately optimal choice may

JEL Classification: D86, G21 and G28
Keywords: compensation design, financial regulation and principal-agent models

Florian Hoffmann
Johann Wolfgang Goethe University
Grüneburgplatz 1
60323 Frankfurt am Main
GERMANY

Email: fhoffmann@finance.uni-frankfurt.de

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=169529

Roman Inderst
Goethe Universität Frankfurt
Grueneburgplatz 1
60323 Frankfurt am Main
GERMANY

Email: inderst@finance.uni-frankfurt.de

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=145025

Marcus Opp
Haas School of Business
University of California, Berkeley
USA

Email: mopp@haas.berkeley.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=169627

*We thank seminar participants at University of Amsterdam, Berkeley, UPF, the 2013 conference at INF/Stockholm, and Stanford and, in particular, Jeffrey Zwiebel and Sandy Grossman for valuable insights. Inderst gratefully acknowledges financial support from the ERC (Advanced Grant "Regulating Retail Finance").

Submitted 15 February 2014

1 Introduction

In the wake of the ongoing financial crisis, 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 backloading 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 even 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 firm to compensate the respective agent at any point in time, conditional on all performance-relevant information that is available until then. Second, we use a multi-task framework, in which the firm 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. This generates the 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 al-

[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.

ready been adopted. In the EU, a new directive adopted in 2010 includes strict rules 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 reflects firms' and society's interests. Apart from the size of compensation packages, it is again the timing of compensation that is at the heart of the debate - in particular, to what extent current practices induce short-termism instead of focusing on long-term performance.[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 tasks, 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 firm. We analyze the determinants of the optimal timing of the long-term bonus and of the weights that are given to the up-front versus the long-term bonus.

Our main contribution is to analyze the implications of mandating a longer deferral of contingent pay. Under such regulation, a bonus must be postponed until a stipulated

[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 insurances' 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).

minimum time or it must be made with the provision that it can still be clawed back until then. Such regulation has essentially two implications. First, any bonus is optimally made contingent also on the performance of the respective deals or of the bank as a whole, rather than being contingent only on the conclusion of a deal or on volume growth. Second, by imposing a minimum deferral period, 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. Motivated by the aforementioned policy discussion, in this paper we are interested in the following question: Will this regulation unambiguously lead to higher diligence? As noted previously, we find that this is not the case, and we identify essentially three effects that a mandatory deferral has on a bank's willingness 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 further delay contingent pay and, thereby, make any bonus contingent also on the respective performance, such as the default of loans, these incentives that previously were targeted exclusively at the acquisition task will now, in addition, induce higher diligence. A second positive effect arises when regulation is particularly restrictive. In this case regulation effectively induces a lower bound on diligence below which the bank's compensation costs are unaffected by diligence. As a result, diligence levels below this threshold are strictly dominated and never implemented. Still, there is also a negative effect of regulation, which even dominates when the bank has itself high incentives to induce diligence or when the second task of acquisition and business growth is of relatively low importance. Then, forcing the bank to delay its bonus will not only increase the level of compensation costs, but it will also make it more rather than less expensive to induce higher diligence, i.e., it increases the *marginal* compensation costs from inducing higher diligence. This is robustly so despite the fact that, by mandating deferral, regulation ensures that the bank can use more information before making a bonus payment. In equilibrium, mandatory deferral of the bonus may still leads, via this effect, to lower rather than higher diligence.

One robust insight of our analysis 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 firm to its agents. When these are high, a binding mandatory deferral of incentive pay will lead to higher equilibrium diligence, while such regulation backfires when acquisition incentives are low. In fact, in the latter case we find that it is optimal not to impose any such restrictions. In terms of observable characteristics, one might conclude that regulation is more likely to have a beneficial impact when the respective agent truly performs the twin tasks of acquiring customers and deals as well as of exerting diligence in concluding or managing this business. Instead, when the institution itself provides sufficient checks or splits tasks entirely, then regulation risks backfiring and leads to less diligence in equilibrium. Also, mandatory deferral risks backfiring when the bank has itself already high incentives to induce diligence, e.g., as the risks that diligence reduces are largely borne by the bank itself, while regulation should increase equilibrium diligence when the provision of acquisition incentives becomes more onerous for the bank as, for instance, competition intensifies.

Literature. 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] According to the theory advanced in Thanassoulis (2012), competition for bankers drives up market levels of remuneration and, thereby, increases banks' default risk. Regulation can react by imposing limits on the proportion of the balance sheet used for bonuses. 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. 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. Inderst and Pfeil (2013) consider compensation regulation in a setting where there is an immediate 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

[6] In fact, in terms of risk-taking, there is 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)).

business opportunities.[7] 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 inherently dynamic nature of our analysis links our paper more broadly to the larger literature that analyzes the optimal mix of short- and long-term compensation for corporate executives, including Peng and Roell (2011), Chaigneau (2012), and Edmans et al. (2012). In this literature, long-term (equity) compensation, in the form of long vesting periods, is considered to effectively link executive compensation to long-term firm performance and, thus, to avoid myopic, short-termist behavior. Still, early vesting may be part of an optimal contract as it allows to reduce compensation risk for risk-averse managers or to smooth their consumption over time.[8] In our model, the manager is risk-neutral and protected by limited liability. However, deferring compensation is still costly as the manager is relatively more impatient. On the other side, deferral increases the informativeness of the performance measure.[9]

What is also different from much of this literature is that the key task of diligence 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, instead, on the implications that regulation has on optimal incentive compensation and, thereby, on the equilibrium provision of diligence.[10]

Organization. The rest of this paper is organized as follows. Section 2 introduces the baseline model. Section 3 derives the optimal compensation contract and equilibrium

[7] More formally, 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" (e.g., Levitt and Snyder (1997) or, more recently, Gromb and Martimort (2007), as well as the applications in Inderst and Ottaviani (2009) and Heider and Inderst (2012)).

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

[9] 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).

[10] In Appendix B we provide a further elaboration of the relationship between our paper and theirs. See also the subsequent remarks after Proposition 1, in particular.

incentive provision without regulation. Section 4 asks whether compensation regulation will induce higher diligence. Section 5 offers some concluding remarks. All proofs are collected in Appendix A, Appendix B contains some additional material.

2 Model

We study a principal-agent problem between a bank and one of its agents, which may be a senior executive, an employee dealing with the bank's customers, or a broker who distributes its products. 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. In Section 4 we study how this principal-agent relationship is affected by regulation of compensation contracts.

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 - possibly relatively infrequent - "bad event" occurs or is avoided. This setting thus encompasses scenarios in which lack of diligence might only be exposed with considerable delay or only in extreme times. We discuss several applications below. 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.

Our focus is on the impact that regulation has on diligence provision in equilibrium. In light of this, we make the following specifications. 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 $c(\mu)$ is twice continuously differentiable. To obtain an interior solution, we stipulate that $c'(\mu) = 0$ for $\mu = 0$, $c'(\mu) > 0$, $c''(\mu) > 0$, and that $c'(\mu)$ becomes sufficiently large as $\mu \to 1$. For tractability we consider two levels of acquisition effort, which is $a_l = 0$ at zero disutility for the agent and $a_h = 1$ at disutility $k > 0$.[11]

[11] The specification for a_l and a_h also allows to interpret our optimal compensation design results in terms of an ex-ante participation constraint. This allows for a closer comparison of our results with those in Hartman-Glaser et al. (2012) and Malamud et al. (2013) (cf. also Appendix B). Note, however, that this analogy of the (latter formalized) acquisition incentive constraint with a participation constraint no

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). This assumption is common in the literatures on labor, executive compensation, and contracting.[12] 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 payments must be non-negative and can be conditioned on 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 $Y_t = 0$ otherwise.[13] Our restriction is to a countable grid of times T_i at which payments b_i are thus paid if and only if the event has not occurred by time T_i, i.e., if $Y_{T_i} = 0$.[14] Note finally that there can be a positive payment b_0 in $T_0 = 0$, i.e., before the bad event can even occur with positive probability.

Gross of compensation costs, when acquisition was successful the bank obtains expected profits $\pi(\mu)$, which is continuously differentiable. For specific applications, one can impose more structure on $\pi(\mu)$, though this is not necessary for our present purpose. For simplicity only, we further set the acquisition effort equal to the likelihood with which acquisition is successful, so that the bank's ex-ante expected profits - still gross of compensation - are given by $a\pi(\mu)$.

Application: Loan Defaults with Externalities. Here, we consider a single consumer loan or mortgage that the agent generates with probability a. The agent may be a broker or an employee of the bank. Through exerting (diligence) effort μ the agent can decrease the likelihood with which a loan subsequently defaults.[15] There may be various

longer holds under regulation. In the latter case, there would not be a contingent (on acquisition) upfront bonus but, instead, possibly a fixed wage, which the considered regulation would not mandate to be deferred.

[12] Cf. Rogerson (1997), DeMarzo and Duffie (1999) or DeMarzo and Sannikov (2006). In the literature, this common assumption is justified on various grounds. For instance, employees may have higher liquidity preferences than the firm does, as they are (more) credit-constrained. In addition, deferred compensation, unless it is securely "ring-fenced," carries the additional risk for the agent that the employer may, through opportunistic behavior, fail to honor his commitment. This generates effectively impatience of the agent.

[13] 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.

[14] As our model is set in continuous time, the restriction is thus to rule out rates at which payments can be made continuously. In Appendix B we appeal to results in Malamud et al. (2013) and argue that such rates are indeed not optimal in our setting with risk neutrality.

[15] This application follows Hartman-Glaser et al. (2012). With the chosen specification, diligence does not impact on the likelihood of making a loan, e.g., through applying a higher standard (as in Inderst and

reasons for why what is privately optimal for the bank may not be welfare optimal. One such reason may be the existence of private bankruptcy costs for consumers and that the bank is more sophisticated when predicting loan default. Moreover, even when there is no such moral hazard problem in the contractual relationship between the bank and the household that takes out a loan, they should both 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.[16] The existence of such externalities from default will motivate our analysis of regulation below.

Application: (Unsuitable) Financial Advice. Take the sale of a long-term savings or investment product or the sale of an insurance. The agent's diligence increases the likelihood that the respective product, which is sold with advice, matches the client's preferences and needs. Whether the respective product is suitable or not may not be evident immediately, e.g., as the burden from the purchase of an illiquid savings or life insurance product that commits a household to paying high contributions or premiums is only felt when the household faces liquidity needs. Likewise, an insurance may fail to provide adequate coverage. In these cases, the "bad event" for the bank or insurance company, as well as for the policy holder, may arise when the latter chooses to cancel the contract or sues for liability. The outcome may be inefficient when there is a commitment problem when wary consumers anticipate a low level of diligence and thus have only a low willingness-to-pay. As argued in Inderst and Ottaviani (2012), consumers may, however, often remain naive about the true conflict of interest when advice is given and

Ottaviani 2009 with an application to consumer financing or in Inderst and Pfeil 2013). This allows to abstract from any built-in tension between the two tasks.

[16] More precisely, we could envisage that default leads to a non-internalized social loss of size X. Then, the expected negative externality for a given level of diligence and given a discount rate r for X is

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

This expected loss is not internalized by the bank and the respective borrower. In case a monopolistic bank could extract all consumer surplus, the surplus under a perpetual loan contract could be captured by some continuous flow payoff F until default. If we normalize the upfront cost of a loan to one and stipulate zero recovery value, we have

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

thus also about the true level of diligence that they should rationally anticipate.[17] This again motivates our subsequent analysis of regulation.[18]

3 Equilibrium without Regulation

3.1 Optimal Compensation without Regulation

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

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

Note 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 respective first-order condition must be satisfied:

$$\sum\nolimits_i b_i e^{-r_A T_i}\left(e^{-\lambda_L T_i} - e^{-\lambda_H T_i}\right) = c'(\mu). \tag{1}$$

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,

[17] For a different take on households' trust in financial advice see Gennaioli et al. (2012).

[18] To be more specific, we could envisage advice on an insurance product. With probability μ the advised product is fully suitable and fully protects the consumer against the specific event ($\lambda_L = 0$). With probability $1-\mu$, instead, the recommended contract was not suitable and leads to an uncovered loss X that arises with arrival rate λ_H. In the case of an uncovered loss, the consumer cancels the insurance contract, which for the firm leads to a future loss of the respective premium, say C with $C > r + \lambda_H$ (where r is the consumer's discount rate). If a consumer expects a level of diligence $\widehat{\mu}$, the consumer's expected utility equals

$$\widehat{\mu}\left(\frac{X-C}{r}\right) - (1-\widehat{\mu})\frac{C}{r+\lambda_H},$$

while given the true μ the firm's expected profits (absent wage costs) equal

$$\pi(\mu) = \frac{C}{r+\lambda_H} + \mu\left(\frac{\lambda(C-r-\lambda_H)}{r(r+\lambda_H)}\right).$$

Welfare is now the sum of the consumer's utility, bank profits, and the agent's expected compensation. A commitment problem arises as the firm could extract a larger premium if it was credible that it induced the agent to exert a higher level of diligence. Note also that in this case it may not be feasible for a non-regulated firm to disclose the agent's total incentives in such way that this is both credible and understood by consumers.

$a_h = 1$, it must hold, in addition, that $V_A(a_h, \mu) \geq V_A(a_l, \mu)$:

$$\sum\nolimits_i b_i e^{-r_A T_i} \left[\mu e^{-\lambda_L T_i} + (1-\mu) e^{-\lambda_H T_i}\right] - c(\mu) \geq k. \tag{2}$$

The total cost of compensation to the bank equals

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

which now uses the bank's (the principal's) discount rate r_P. In what follows, it will be convenient to use $\Delta_r = r_A - r_P > 0$, 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 rates with which a bad event occurs, it captures the speed of learning.[19] 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 (1) and (2) 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 (2), but not the diligence incentive constraint (1). 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 (1) this must satisfy

$$b_T = c'(\mu) \frac{e^{(r_A + \lambda_H)T}}{e^{\Delta_\lambda T} - 1}. \tag{3}$$

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 (3), and a unique timing T. If*

$$\mu < \frac{1}{2}\left(1 - \frac{\Delta_r}{\Delta_\lambda}\right) \tag{4}$$

holds and the costs of acquisition effort k are sufficiently large, an additional up-front bonus $b_0 > 0$ is paid:

$$b_0 = k + c(\mu) - c'(\mu)\left(\mu + \frac{1}{e^{\Delta_\lambda T} - 1}\right). \tag{5}$$

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

Proof. Follows from Proposition 4 below.

We now comment on this characterization.[20] For this we take first the key condition (4), 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 this case, actually the bank's objective 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.[21] 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 (4) 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, which is adjusted so as to preserve the agent's incentive to choose a given μ.

When condition (4) does not hold, deadweight loss from delay would always become strictly lower as T decreases.[22] 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] Hartman-Glaser et al. (2012) analyze a similar model with binary diligence effort. However, they impose a parameter restriction that essentially implies that (4) does not hold, so that an up-front payment is never optimal (cf. also the discussion in Appendix B).

[21] Precisely, as we presently consider as fixed the induced level of diligence, together with acquisition effort $a_h = 1$, 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. After substituting from (3), we have

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

which is indeed zero when $\Delta_r = 0$ as both parties are equally impatient, $r_A = r_P$.

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

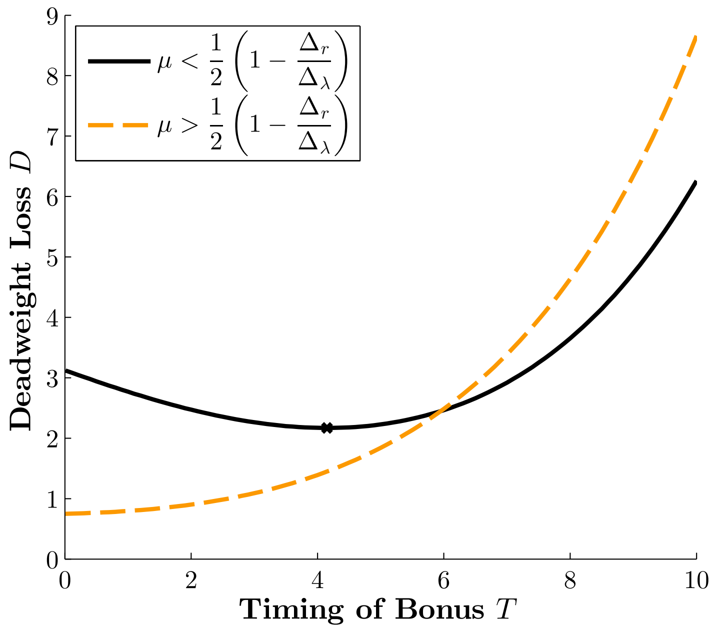

Figure 1. This graph plots "deadweight loss" as a function of the timing of the long-term bonus T. The solid line plots the case in which an interior minimum obtains, i.e., $\mu = 0.05 < \frac{1}{2}\left(1 - \frac{\Delta_r}{\Delta_\lambda}\right)$ and $c(\mu) = 25\mu^2$. The dotted line plots the case 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 (3), 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 (5), which is the up-front bonus that is paid additionally to induce high acquisition effort.

We continue by providing additional details for the characterization obtained in Proposition 1. These follow as well from the proof of Proposition 4 below.

Corollary 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 (2) is slack (Regime 1) and T is given by*

$$T_1 = \frac{1}{\Delta_\lambda} \ln \left(1 + \frac{\Delta_\lambda - \Delta_r + \sqrt{(\Delta_\lambda - \Delta_r)^2 + 4\Delta_\lambda \Delta_r \mu}}{2\Delta_r \mu} \right). \tag{6}$$

For $k \geq \overline{k}$, 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

$$T_2 = \frac{1}{\Delta_\lambda} \ln \left(1 + \frac{c'(\mu)}{k + c(\mu) - c'(\mu)\,\mu} \right). \tag{7}$$

This applies when either condition (4) does not hold or always when k is still sufficiently low with $k \leq \overline{k}$. When, instead, (4) holds and $k > \overline{k}$, Regime 3 applies with $b_0 > 0$ and T_3 as the unique positive solution for T in

$$\frac{1 - e^{-\Delta_r T}}{1 - e^{-\Delta_\lambda T}} \frac{1}{1 + \mu \left(e^{\Delta_\lambda T} - 1 \right)} = \frac{\Delta_r}{\Delta_\lambda}. \tag{8}$$

The thresholds on acquisition costs satisfy:

$$\underline{k} = c'(\mu) \left(\mu + \frac{1}{e^{\Delta_\lambda T_1} - 1} \right) - c(\mu), \tag{9}$$

$$\overline{k} = c'(\mu) \left(\mu + \frac{1}{e^{\Delta_\lambda T_3} - 1} \right) - c(\mu) \text{ for } \mu < \frac{1}{2} \left(1 - \frac{\Delta_r}{\Delta_\lambda} \right). \tag{10}$$

Further Discussion and Comparative Analysis. The further characterization of the delay of the long-term bonus in Corollary 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. For regime 3 we also have to consider the fact that there two payments are made. We have from (5) that the up-front bonus b_0 increases one-for-one with k, while b_T remains unchanged.

Corollary 2 *Holding μ fixed, consider an increase in the costs of acquisition effort k. Then, over all regimes compensation becomes more "front-loaded" in the following sense:*

The delay of the long-term bonus T (weakly) decreases and, when this is paid, the up-front bonus b_0 strictly increases while then the long-term bonus b_T remains unchanged.

Corollary 2 thus captures intuitively the notion that when the agent must receive high incentives to induce acquisition effort, this renders his compensation more front-loaded. At the opposite extreme, when from $k = 0$ acquisition effort must not be induced at all, the agent will never receive an up-front bonus and his long-term bonus is paid relatively late.

The characterization of regimes in Proposition 1 and Corollary 1 is in terms of acquisition cost k. Recall that presently the choice of diligence μ that the bank wants to implement is still exogenous. To complete the characterization, we now complement Proposition 1 with a 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 Corollary 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 3 *Consider a given level of acquisition cost $k > 0$. Then, the following regimes from Proposition 1 apply, depending on the bank's choice of implemented level of diligence μ: Regime 3 (with $b_0 > 0$) applies when μ is low (provided that $\Delta_r < \Delta_\lambda$), regime 2 (with binding acquisition constraint but $b_0 = 0$) applies for intermediate levels, and regime 1 (with slack acquisition constraint) 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. 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.

Finally, we can determine the comparative statics of the optimal bonus times in μ using the different regimes described in Corollary 3.

Corollary 4 *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, 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.*

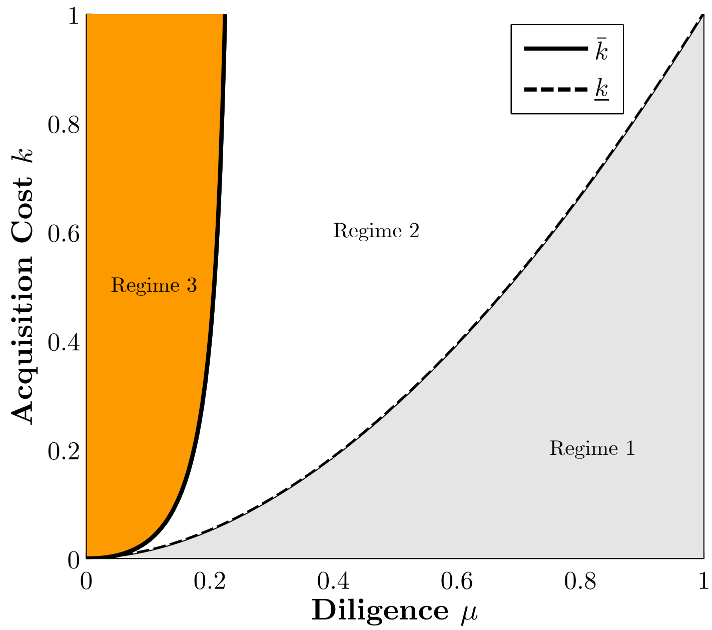

Figure 2. This graph plots the (μ, k) combinations that give rise to the three regimes described in Proposition 1. The cost function for diligence satisfies $c(\mu) = 0.5\mu^2$. The remaining parameters are $\Delta_r = 0.3$, $\Delta_\lambda = 0.6$.

Proof. See Appendix A.

Corollary 4 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, Figure 3 depicts the equilibrium choice of delay T as a function of the implemented level of diligence μ.

3.2 Equilibrium Provision of Diligence

From the characterization of the optimal contract in Proposition 1 we can obtain for any given μ the minimum compensation costs. We denote this 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

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

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

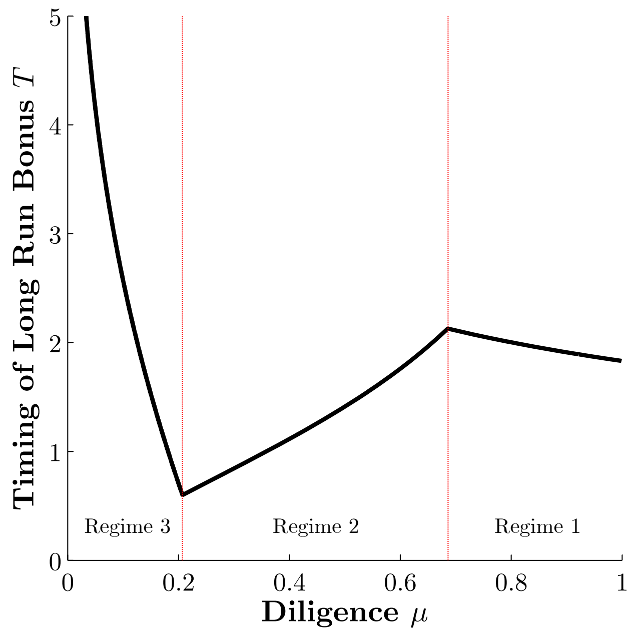

Figure 3. This graph plots the equilibrium choice of delay T as a function of the implemented level of diligence μ. As in Figure 2 the cost function for diligence satisfies $c(\mu) = 0.5\mu^2$. The remaining parameters are $\Delta_r = 0.3$ and $\Delta_\lambda = 0.6$. The cutoffs for the respective regimes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

condition

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

Note that without additional restrictions on functional forms, μ^* may not be pinned down uniquely such that μ^* may be set-valued.[23] 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 Corollary 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 (4) holds so that indeed $b_0^* > 0$ for high k.

[23] 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.

Proposition 2 *When the implemented level of diligence μ^* is optimally chosen by the bank, we have the following characterization result, making use of the three regimes introduced in Proposition 1 and Corollary 2. For $k < \underline{k}^*$ regime 1 applies; for $k > \overline{k}^*$ and and when Δ_r/Δ_λ is sufficiently small, regime 3 applies; otherwise, regime 2 applies.*

Proof. See Appendix A.

4 Regulation of Compensation

In this section, we consider the impact of regulation. As noted in the introduction, we are interested in an analysis of a particular, frequently discussed and even implemented regulation that requires all bonus compensation to be paid out after a certain time $\tau > 0$. It is evident that equilibrium compensation contracts then endogenously require all pay to be contingent on both acquisition as well as subsequent performance, i.e., absence of the "bad event." This follows directly from risk neutrality: The agent's acquisition constraint only depends on the expected *level* of pay while conditioning on all available information at the time of the payout additionally provides incentives to exert higher diligence. For expositional reasons, it is however useful to specify that government regulation *"requires"* payments to be conditional on all information at the time of payout.[24]

We analyze whether through such regulation diligence can be increased, thereby reducing, for instance, the potential for excessive risk-taking or unsuitable advice. Formally, we thus ask in what follows how the equilibrium level of diligence μ^* changes under such regulation and, more specifically, how it changes when the respective mandatory deferral period τ is adjusted. As noted above, there are various reasons, depending on the particular application, for why the unregulated equilibrium outcome leads to an inefficiently low level of diligence. Still, we postpone a broader discussion of welfare and policy until the end of this section.

Compensation regulation targets the agents' principal, namely the bank in our applications. The question must thus be whether mandatory deferral induces the bank to restructure the agent's incentives accordingly. For this, we first characterize in section 4.1 the optimal compensation choice under regulation. Subsequently, in section 4.2 we

[24] This allows us to ignore the analysis of compensation contracts for *dominated* levels of diligence. A diligence level $\tilde{\mu}$ is *dominated* if a higher level of diligence $\mu > \tilde{\mu}$ can be implemented at the same expected cost to the principal.

ask what level of diligence the bank wants to optimally induce under regulation. Our procedure thus mirrors the steps of the analysis without regulation. Formally, regulation mandates that the timing of any bonus b_i must now satisfy $T_i \geq \tau$. Note that, in particular, this rules out an up-front payment, b_0, as this conditions only on whether a deal was made or not.[25]

4.1 Characterization of the Compensation Contract with Regulation

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 complicates the exposition. As a consequence, we proceed stepwise.

The first thing to note is that now a given level of diligence μ 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 a single bonus is infinitely 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, becomes sufficiently large.[26]

Proposition 3 *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 (7). Further, the maximum feasible delay, $\tau(\mu) = T_2$, is strictly decreasing in k and strictly increasing in μ.*

Proof. See Appendix A.

[25] This would clearly be different when the optimal compensation included an unconditional initial payment, corresponding to a fixed wage. This is, however, not optimal in our model. Incidentally, however, note that such a fixed wage could arise when the acquisition constraint was interpreted as a participation constraint (with k as the agent's reservation utility). Then, in regime 3, b_0 would represent a fixed wage that could still be paid under the considered regulation.

[26] Note again, that we presently look at the auxiliary problem where regulation *requires* payments to be conditional on all information at the time of payout, giving rise to the non-implementability result for low levels of μ for sufficiently high τ. While these diligence levels are implementable under the originally considered policy, where regulation only prescribes payments to occur after τ, they are *dominated* in the sense of footnote 24 and will, hence, never be chosen in equilibrium.

We may also rewrite the implementability restriction of Proposition 3 as follows. As the threshold characterized in Proposition 3, $\tau(\mu)$, is strictly increasing in μ, the inverse function exists. This obtains, now for given τ, a unique threshold $\underline{\mu}(\tau) > 0$, such that only diligence levels $\mu \geq \underline{\mu}(\tau)$ are compatible with the respective minimum deferral time $\tau > 0$. The minimum diligence level $\underline{\mu}(\tau)$ that this requires is a strictly increasing function of τ. This result already points to a rather immediate effect of regulation, which we further discuss below.

Turn now to the case where for given diligence level μ, together with values k and τ, it holds that $\tau \leq T_2$. Hence, it is then indeed feasible for the bank to incentivize μ while adhering to the mandatory deferral imposed by regulation. The further characterization follows intuitively the same principles as Proposition 1 and Corollary 1.

Proposition 4 *Take the case where a given level of diligence μ, next to acquisition effort, can be implemented under regulation (as $\tau \leq T_2$). Then, the cost-minimizing compensation contract again consists of at most two payments, which now occur at τ and/or at a uniquely determined $T > \tau$. Precisely, depending on the costs of acquisition effort k, the following characterization obtains:*

i) If $k < \underline{k}$, with $\underline{k}$ given by (9), the acquisition constraint is slack (regime 1) and there is a single payment b_T satisfying (3). The optimal delay of the long-term bonus is uniquely determined from $T = \max\{T_1, \tau\}$, where T_1 is given by (6).

ii) If $k \geq \underline{k}$, there are two subcases to consider: If the following condition holds:

$$\frac{\Delta_r}{\Delta_\lambda} < \frac{1 - \mu\left(1 + e^{\Delta_\lambda \tau}\right)}{1 + \mu\left(e^{\Delta_\lambda \tau} - 1\right)}, \tag{11}$$

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 (1) and (2) (regime 3). Consequently, the optimal payout times are τ and $T = T_3(\tau) > \tau$, which is the unique solution $T > \tau$ to

$$\frac{1 - e^{-\Delta_r(T-\tau)}}{1 - e^{-\Delta_\lambda(T-\tau)}} \frac{1 + \mu\left(e^{\Delta_\lambda \tau} - 1\right)}{1 + \mu\left(e^{\Delta_\lambda T} - 1\right)} = \frac{\Delta_r}{\Delta_\lambda}. \tag{12}$$

If either (11) is violated or $k \leq \overline{k}(\tau)$, then there is again a single payment b_T satisfying (3), which now occurs at $T = T_2$, as defined in (7) (regime 2).

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 two long-term bonus payments: one at τ and one strictly later at some time $T > \tau$. (This corresponds to the first subcase in assertion ii) of Proposition 4.) We can show in this case that when regulation becomes gradually more severe as τ increases, $T = T_3(\tau)$ 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 (11) 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 (4) 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.

4.2 Comparative Analysis in the Mandatory Time of Deferral

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 4. However, now the bank faces the additional constraint that $T \geq \tau$ for any bonus, which - as obtained in Proposition 3 - also constrains the feasible set of diligence levels that the bank can induce: $\mu \geq \underline{\mu}(\tau)$. Recall that the implementability threshold $\underline{\mu}(\tau)$ is strictly increasing in τ with $\underline{\mu}(0) = 0$. In the following analysis, it will also be useful to analyze the effect on this diligence constraint when the mandatory deferral time becomes extremely long.[27] Formally, we define $\overline{\mu} = \lim_{\tau\to\infty} \underline{\mu}(\tau)$ which is implicitly characterized by:

$$k + c(\overline{\mu}) - c'(\overline{\mu})\overline{\mu} = 0. \tag{13}$$

Intuitively, $\overline{\mu}$ is strictly increasing in the respective cost of acquisition effort k.[28] 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. Further, as these costs k become arbitrarily close to zero with $k \to 0$, we have also $\overline{\mu} \to 0$ (albeit this should not suggest that it is then optimal for the bank to induce such a low level of diligence).

In what follows, 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.[29] Again, note that as in the case without regulation, the now constrained problem may not have a unique solution.

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 not bind (namely in regime 3).

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

Proposition 5 *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 exists a threshold for the minimum deferral period $\widetilde{\tau} > T_1$ so that equilibrium diligence $\mu^*(\tau)$ is strictly decreasing in τ for $\tau < \widetilde{\tau}$ and strictly increasing for $\tau > \widetilde{\tau}$.*

[27] 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.

[28] Formally, this follows directly from the implicit function theorem and the properties of $c(\mu)$.

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

Proof. See Appendix A.

The results of Proposition 5 are due to the interaction of two effects, which we explain next. Regulation increases the *marginal* compensation costs to induce higher diligence effort.[30] This effect is responsible for the decrease in the optimally induced diligence effort as τ increases, at least when τ is not yet too large. Hence, when regulation indeed constrains the bank's optimal choice as $\tau > T_1^*$, it increases not only the total costs of compensation, as we discuss in more detail below, but as τ increases, it becomes for the bank increasingly expensive to induce (marginally) higher diligence. Rather than inducing the bank to implement higher diligence when it has to wait longer and can thus learn more before a bonus is paid, the imposition of a mandatory deferral period then leads to a strictly lower equilibrium level of diligence.

This negative effect of regulation is, however, counteracted by a positive effect. Recall once more that the agent's payoff, when $b_0 = 0$, is for given μ decreasing in the time of compensation. (Put differently, the agent's rent decreases as the bonus is further delayed.) When the bank is required to postpone compensation always further, this increases the minimum level $\underline{\mu}(\tau)$ that the bank *has* to implement in order to still satisfy the acquisition constraint. From a certain threshold $\widetilde{\tau}$ onwards, this additional constraint binds and from thereon it is optimal for the bank to set $\mu^*(\tau) = \underline{\mu}(\tau)$, so that the constrained optimal choice is just equal to the lowest diligence level that is compatible with regulation and with the acquisition constraint. 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. The respective cases for when equilibrium diligence remains constant, decreases, or increases in the minimum deferral time are depicted in the illustration in Figure 4. In this example, equilibrium diligence is highest when no (binding) regulation is imposed. Clearly, this is always the case when the equilibrium level of diligence without regulation in regime 1 is strictly above $\overline{\mu}$, which is the limit of the boundary $\underline{\mu}(\tau)$ as $\tau \to \infty$. Whether the unregulated outcome thus generates the highest level of diligence in the presently analyzed case of regime 1 or not depends thus on the bank's own incentives to incentivize high diligence. We return to this later when we summarize the overall impact that regulation has on diligence.

[30] Formally, with W_1 as the respective compensation costs in regime 1, we have for $\tau \geq T_1$ that $\frac{d}{d\tau}\left(\frac{dW_1}{d\mu}\right) > 0$.

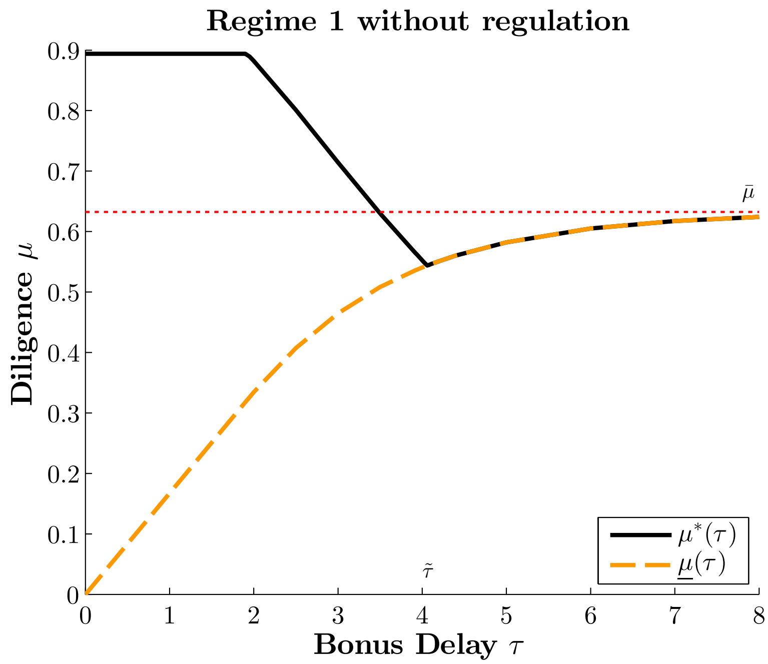

Figure 4. This graph plots the equilibrium diligence μ^* as a function of the mandatory delay of compensation τ for the case when regime 1 obtains without regulation. For small τ, the regulatory constraint does not bind and hence $\mu^* = \mu_1^*$. Once the constraint binds, equilibrium diligence is 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 $\underline{\mu}(\tau)$. As $\mu_1^* > \bar{\mu}$, case i) of Proposition 5 obtains. 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 $r_A = 0.2$.

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

Proposition 6 *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 6. First, for regime 2 recall that the regulatory constraint becomes binding 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 (see proof of Proposition 6), 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.

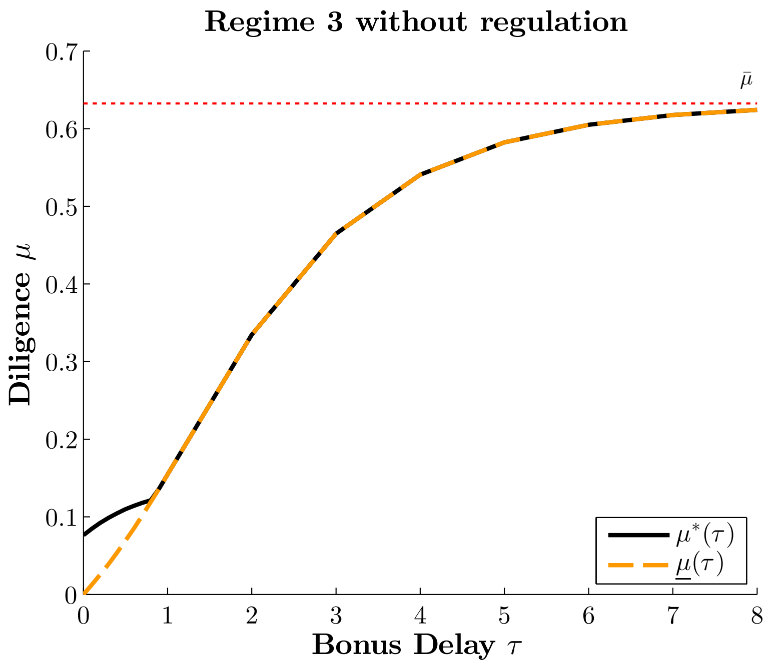

Figure 5. This graph plots the equilibrium diligence μ^* as a function of the mandatory delay of compensation τ for the case when regime 3 obtains without regulation. Since an upfront bonus is paid without regulation, regulation has an immediate positive effect on equilibrium diligence. Once the constraint $b_T \geq 0$ 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.

4.3 Impact of Deferral Regulation

Combining Propositions 5 and 6, we can immediately make the following observations. Surely, when $k \geq \underline{k}^*$ (regimes 2 and 3 without regulation) 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 boundary 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 $\overline{\tau}$.[31]

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 does so. 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 the impact that mandatory deferral has on equilibrium diligence.

Proposition 7 *There exists a cutoff on the costs of acquisition effort, so that for low values of k the highest equilibrium diligence $\mu^*(\tau)$ is achieved when no binding deferral regulation is imposed. Instead, for all higher values of k equilibrium diligence is highest when the minimum deferral period τ is made as high as possible, in particular equal to $\tau = \overline{\tau}$ in case the bank's zero-profit constraint must be satisfied.*

Proof. See Appendix A.

[31] While this paper only considers binary acquisition effort, we conjecture that, when allowing for acquisition effort to be a continuous choice, 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.

When k is relatively high, this may suggest that deal or customer acquisition is a main part of an agent's work. Proposition 7 suggests that imposing mandatory deferral is, instead, likely to backfire when this is not the case, e.g., as 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.") Depending on the respective applications, such as a particular financial product or sales channel, regulation should and could possibly be applied differently, e.g., through the creation of a "safe haven" when sufficient provisions are made that effectively limit the extent to which the agent will act like a salesperson by prospecting for new business.

Alternatively, a 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 Proposition 7 would suggest that regulation of deferred incentive pay could be (more) beneficial when competition intensifies. Recall finally that for regime 1 the threshold on k in Proposition 7 depends on a comparison with the equilibrium outcome without regulation, $\mu^*(\tau = 0)$: The higher is this value, the higher is this threshold on k, so that the range of parameters increases for which any binding mandatory deferral leads to lower rather than higher diligence. There may be various reasons for why the bank may have high incentives - as captured by the function $\pi(\mu)$ - to induce more diligence, and these reasons should depend on the particular application. For instance, with loans the bank arguably cares more about diligence when it keeps a larger fraction on its own books, while in the case of suitable advice stricter legal enforcement of liability should also raise the bank's incentives accordingly.[32] The choice of a particular application and, thereby, of $\pi(\mu)$ together with a specification of the externality would also allow to go beyond Proposition 7 in an analysis of the optimal choice of deferral period for the case when a longer deferral time can indeed lead to higher diligence.[33]

[32] Note that such changes would affect the function $\pi(\mu)$, while not the primitives used for our characterization, such as Propositions 1 and 4 (i.e., k, Δ_r, and Δ_λ).

[33] In particular, note that even when we could ignore the bank's zero-profit constraint, then also for k above the threshold in Proposition 7 there would always be a bounded value of τ that maximizes welfare, i.e., the sum of the bank's and the agent's expected payoff minus the externality (cf. the applications in Section 2). This holds as for $k > 0$ the deadweight loss increases without bounds as $\tau \to \infty$.

5 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 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. These predictions could now be sharpened by applying our model in a particular context, such as that of the two applications that we sketched. In a particular application, such as to (non-)suitable advice or loans, the relative importance of the acquisition and the diligence task as well as the costs of inducing the respective effort could then be linked to primitives such as the way the considered products are bought or sold. Likewise, the need for regulation could then be modeled explicitly and would determine the importance of diligence effort from a welfare perspective.

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. Clearly, when the regulator knows all parameters and has the power to do so, he could simply dictate a particular compensation contract - or even prescribe for the principal (bank) a sufficiently large penalty in case the bad event occurs. Future work could turn to the question when compensation regulation should optimally be part of such a more broadly designed regulation.

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. 2010. Paying for Long-Term Performance. *University of Pennsylvania Law Review* 158, 1915–60.

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, 2108–31.

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, 2681–724.

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.

Gennaioli, N., Shleifer, A., and R.W. Vishny. 2012. Money Doctors. NBER Paper 18174.

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, 2381–415.

Inderst, R., and M. Ottaviani. 2009. Misselling Through Agents. *American Economic Review* 99, 883–908.

Inderst, R., and M. Ottaviani. 2012. How (Not) to Pay for Advice: A Framework for Consumer Financial Protection. *Journal of Financial Economics* 105, 393–411.

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

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, 641–61.

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.

6 Appendix A: Omitted Proofs

Proof of Corollary 4. When T is determined by (6), 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 (7), 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 (8), 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 4.[34] This further implies that b_0, as defined by (5), 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.

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.) As the respective solution may not be unique, define the highest such value by $\overline{\mu}_i^*$ and the respective lowest solution by $\underline{\mu}_i^*$, $i\in\{1,2,3\}$. The respective contractual parameters are indexed accordingly. Define then, in complete analogy to the thresholds in (9) and (10),

$$\underline{k}^*=c'\left(\overline{\mu}_1^*\right)\left(\overline{\mu}_1^*+\frac{1}{e^{\Delta_\lambda T_1(\overline{\mu}_1^*)}-1}\right)-c\left(\overline{\mu}_1^*\right),$$
$$\overline{k}^*=c'\left(\underline{\mu}_3^*\right)\left(\underline{\mu}_3^*+\frac{1}{e^{\Delta_\lambda T_3(\underline{\mu}_3^*)}-1}\right)-c\left(\underline{\mu}_3^*\right).$$

[34] In particular, Lemma A3 below shows that, when $k>\overline{k}$ and (4) holds, then $f(T)$ has a unique interior non-zero T^* and is sloping upwards at T^*.

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

$$\underline{\mu}_3^* < \frac{1}{2}(1 - \frac{\Delta_r}{\Delta_\lambda}). \tag{14}$$

We now turn to the proof of the proposition.

The equilibrium wage cost function is given by:

$$W(\mu) = \begin{cases} c(\mu) + k + c'(\mu)\left(e^{\Delta_r T_3(\mu)} - 1\right)\left[\mu + \frac{1}{e^{\Delta_\lambda T_3(\mu)}-1}\right] & \text{for} \quad \mu < \overline{k}^{-1}(\mu) \\ \left(1 + \frac{c'(\mu)}{k+c(\mu)-c'(\mu)\mu}\right)^{\frac{\Delta_r}{\Delta_\lambda}}[k + c(\mu)] & \text{for} \quad \overline{k}^{-1}(\mu) < \mu < \underline{k}^{-1}(\mu) \\ c'(\mu)\, e^{\Delta_r T_1(\mu)}\left[\mu + \frac{1}{e^{\Delta_\lambda T_1(\mu)}-1}\right] & \text{for} \quad \mu > \underline{k}^{-1}(\mu) \end{cases}, \tag{15}$$

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:

$$\mu_1^* = \arg\max\{\pi(\mu) - W_1(\mu)\}, \text{ s.t. } 0 \leq \mu < 1, \tag{16}$$

$$\mu_3^* = \arg\max\{\pi(\mu) - W_3(\mu)\}, \text{ s.t. } 0 \leq \mu \leq \tilde{\mu} := \frac{1}{2}\left(1 - \frac{\Delta_r}{\Delta_\lambda}\right), \tag{17}$$

where $W_i(\mu)$ refers to the wage cost function $W(\mu)$ in regime i (see 15). Note that both problems (16) and (17) 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. Any solution to (16) must be interior[35] and 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 $\tilde{\mu} = \frac{1}{2}\left(1 - \frac{\Delta_r}{\Delta_\lambda}\right)$. 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

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

[35] This follows from the assumptions on $c(\mu)$ and $\pi(\mu)$.

Lemma A1. $\mu_1^* < \mu_3^*$.

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. The result then follows from standard monotone comparative statics results.[36] **Q.E.D.**

Lemma A2. $\underline{k}^* < \overline{k}^*$.

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 $\overline{k}^* = \tilde{k}\left(\underline{\mu}_3^*, T_3\left(\underline{\mu}_3^*\right)\right) > \underline{k}^* = \tilde{k}\left(\overline{\mu}_1^*, T_1\left(\overline{\mu}_1^*\right)\right)$.
Q.E.D.

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 $k \geq \underline{k}^*$ 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

[36] Precisely, we may consider an objective function $\Phi(\mu, \theta) = \theta \Pi_1(\mu) + (1 - \theta)\Pi_3(\mu)$, which has the cross-derivative $d\Phi^2/(d\mu d\theta) < 0$ (i.e., it is single-crossing in (μ, θ)).

regime 2 obtains for any $k \geq \underline{k}^*$. 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 Propositions 3. 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 (1) and (2) 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 (7) and negative for $\tau > T_2$. Hence, the two constraints (1) and (2) 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 (7), and the positive dependence on μ from Corollary 4**. Q.E.D.**

Proof of Propositions 1 and 4. It is convenient to restate the full program, where we take as given that the firm wants to implement μ as well as $a = a_h = 1$:

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

$s.t.$

$$\sum\nolimits_i b_i e^{-r_A T_i}\left(e^{-(\lambda_H-\Delta_\lambda)T_i} - e^{-\lambda_H T_i}\right) = c'(\mu), \tag{18}$$

$$\sum\nolimits_i b_i e^{-r_A T_i}\left[\mu e^{-(\lambda_H-\Delta_\lambda)T_i} + (1-\mu)e^{-\lambda_H T_i}\right] - c(\mu) \geq k, \tag{19}$$

$$T_i \geq \tau,$$

$$b_i \geq 0.$$

Define κ_μ (for (18)), κ_a (for (19)), 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[37]

$$\kappa_a = e^{\Delta_r \tau} - \frac{e^{(r_A+\lambda)\tau}\kappa_{b_\tau} + \kappa_\mu\left(e^{\Delta_\lambda \tau} - 1\right)}{1 + \mu\left(e^{\Delta_\lambda \tau} - 1\right)}. \tag{20}$$

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

$$\kappa_\mu = \left(1 + \left(e^{\Delta_\lambda T_i} - 1\right)\mu\right)\frac{e^{(r_A+\lambda)\tau}\kappa_{b_\tau} + \left(e^{\Delta_r T_i} - e^{\Delta_r \tau}\right)(1-\mu) + \left(e^{\Delta_\lambda \tau + \Delta_r T_i} - e^{(\Delta_r+\Delta_\lambda)\tau}\right)\mu}{e^{\Delta_\lambda T_i} - e^{\Delta_\lambda \tau}}. \tag{21}$$

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 $\kappa_{b_i} = 0$, substituting from (20) and (21) for κ_a and κ_μ, any $T_i > \tau$ must satisfy

$$\begin{aligned}&\Delta_r\left(1 - e^{-\Delta_\lambda(T_i-\tau)}\right)\left(1 + \mu\left(e^{\Delta_\lambda T_i} - 1\right)\right) - \Delta_\lambda\left(1 - e^{-\Delta_r(T_i-\tau)}\right)\left(1 + \mu\left(e^{\Delta_\lambda \tau} - 1\right)\right) \\ &= \Delta_\lambda e^{(r_A+\lambda)\tau - \Delta_r T_i}\kappa_{b_\tau}.\end{aligned} \tag{22}$$

Lemma A3. *Consider equation (22) 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 (22). 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 (22) respectively:

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

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

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

[37] 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 $0 \leq \kappa_{b_0} \leq 1$.

Depending on whether $\partial^2 f\left(T,\tau\right)/\partial T^2$ 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 $\lim_{T\to\infty} f\left(T,\tau\right) = \infty$ there exists a unique solution $T > \tau$ to (22) if and only if $\kappa_{b_\tau} > 0$. Note that

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

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

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

gives

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

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 A3 unique solution $T > 0$ must satisfy (8). The associated payment b_T then follows from (18) and is given by (3). Finally, (20) together with $\kappa_{b_0} = 0$ then imply that $\kappa_a = 1$. Hence, (19) must hold with equality so that b_0 is given by (5). 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 (10) is obtained from setting $b_0 = 0$ in (5).

Case $\tau = 0$, $\kappa_{b_0} = 1$. Then, (20) implies $\kappa_a = 0$, such that the constraint (19) is slack. The first-order condition (22) simplifies to

$$\Delta_r \left(1 - e^{-\Delta_\lambda T}\right) + \Delta_r \mu \left(e^{\Delta_\lambda T} + e^{-\Delta_\lambda T} - 2\right) - \Delta_\lambda = 0, \tag{23}$$

which has a unique solution given by (6), while again b_T is given by (3) and now $b_0 = 0$. This case applies if and only if $k < \underline{k}$, where $\underline{k}$ as given in (9) is obtained from the slack constraint (19). We finally show that $\underline{k} < \overline{k}$, which is equivalent to showing that $T = T'$ solving (23) is larger than $T = T''$ solving (8). This follows as the left-hand-side in (23) 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 (20), the constraint (19) binds. We now obtain from (22), (18), and (19) explicit solutions for T and b_T which are given by (7) and

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

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 4, 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 (6). 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 (23)). The unique payment is then given by (3) with $T = T_1$ for $\tau \leq T_1$ and $T = \tau$ for $\tau > T_1$.[38]

Case $\tau > 0$ and binding acquisition constraint. Take now the case where $k \geq \underline{k}$. When (11) is violated and, hence, from Lemma A3 there is only a single payment, the two binding constraints (18) and (19) imply that this occurs at $T = T_2$ as defined in (7) and is given by (3). Hence, from Proposition 3, 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 (11) holds there can be two positive payments, which are determined from the binding constraints (18) and (19):

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

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

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 (12) satisfies $T_3(\tau) > T_2$.

[38] 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 (19) remains slack only as long as $\tau \leq T_2$, i.e., as long as μ remains implementable (cf. Proposition 3).

Hence, there are two positive payments b_τ and b_T, if and only if μ is implementable, (11) holds and $k > \overline{k}(\tau) := \left[\mu + \frac{1}{\left(e^{\Delta_\lambda T_3(\tau)}-1\right)}\right] c'(\mu) - c(\mu)$. 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 A3). The relevant part of $\frac{\partial f}{\partial \tau}$ then is

$$sgn\left(\frac{\partial f}{\partial \tau}\right) = sgn\left(\Delta_r\left(1-\mu\right)\left(e^{\Delta_\lambda(T-\tau)} - e^{\Delta_r(T-\tau)}\right) + \left(\Delta_r + \Delta_\lambda\right)\mu e^{\Delta_\lambda T}\left(1 - e^{\Delta_r(T-\tau)}\right)\right). \tag{24}$$

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 22) and substituting this into (24) yields

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

$$\begin{aligned}
J_1 &= {\Delta_r}^2\left(e^{\Delta_\lambda(T-\tau)} + e^{-\Delta_\lambda(T-\tau)} - 2\right) - \Delta_\lambda^2\left(e^{\Delta_r(T-\tau)} + e^{-\Delta_r(T-\tau)} - 2\right) > 0,\\
J_2 &= \Delta_r\left(e^{\Delta_\lambda(T-\tau)} - 1\right) - \Delta_\lambda\left(1 - e^{-\Delta_r(T-\tau)}\right) > 0,\\
J_3 &= \Delta_\lambda\left(1 - e^{-\Delta_r(T-\tau)}\right) - \Delta_r\left(1 - e^{-\Delta_\lambda(T-\tau)}\right) > 0.
\end{aligned}$$

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 Proposition 5. With $\tau > 0$, the optimal level of diligence effort for regime 1 solves the following program:

$$\begin{aligned}
&\max_{\mu}\left\{\pi(\mu) - W_1(\mu, t(\mu))\right\}\\
&\qquad s.t.\, t\left(\mu\right) \geq \tau\\
&\qquad \mu \geq \underline{\mu}(\tau),
\end{aligned} \tag{25}$$

where the lower bound on μ is uniquely determined from

$$k + c\left(\mu\right) - \mu c'(\mu) = \frac{c'\left(\mu\right)}{e^{\Delta_\lambda \tau} - 1}, \tag{26}$$

if this admits a positive solution, while we set $\underline{\mu}(\tau) = 0$ else.[39] From Propositions 1 and 4, we further have that $t(\mu) = T_1(\mu)$ as determined from (6) 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

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

where we have used that $\partial W_1 / \partial t > 0$ for $t > T_1$ (cf. the proof of Proposition 1**).** Hence, it follows from standard monotone comparative statics results, that $\mu_1^*(\tau)$ must be decreasing in τ for $\tau \geq T_1(\overline{\mu}_1^*(0))$ as long as $\overline{\mu}_1^*(\tau) > \underline{\mu}(\tau)$.[40] 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 $\overline{\mu}_1^*(\widetilde{\tau}) = \underline{\mu}(\widetilde{\tau})$. Then, for $\tau \geq \widetilde{\tau}$, the optimal level of diligence effort is uniquely determined from $\mu_1^*(\tau) = \underline{\mu}(\tau)$ and increasing in τ up to $\overline{\mu}$, which, thus, is the maximal value of diligence that can be achieved with any $\tau > T_1(\overline{\mu}_1^*(0))$. **Q.E.D.**

Proof of Proposition 6. 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 (4) is violated, or, that $\underline{k}^* \leq k \leq \overline{k}^*$ and (4) holds. Then, the optimal level of diligence $\mu_2^*(\tau)$ is still determined from program (25), where now we have $t(\mu) = T_2(\mu)$ as given by (7) as long as $T_2(\mu) \leq \tau$ and $t(\mu) = \tau$ else. We will call the deferral regulation "binding", when it constrains the bank in the sense of requiring a suboptimal combination of payment date T and diligence μ. Note that from Corollary 4, we have $dT_2/d\mu > 0$, so that, as τ increases, the constraint $T_2(\mu) \geq \tau$ will be violated for $\underline{\mu}_2^*(0)$ first. Next, we will show that, for $\tau > T_2^*(\overline{\mu}_2^*(0))$, where deferral regulation "binds", we have $\mu_2^*(\tau) = \underline{\mu}(\tau)$, which is increasing in τ. To see this assume to the contrary that $\overline{\mu}_2^*(\tau) > \underline{\mu}(\tau)$, for some $\tau > T_2^*(\overline{\mu}_2^*(0))$. This implies, however, that the acquisition constraint is slack such that we are back to regime 1. Then

[39] 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 (26) as long as this is non-negative. Existence of a positive solution to (26) is guaranteed for $\tau > \tau'$, where $\tau' = \frac{1}{\Delta_\lambda} \ln \left(1 + \frac{c'(0)}{k}\right)$ stays bounded, as long as $k > 0$.

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

the arguments in the proof of Proposition 5 above, together with the observation that, by definition, $\overline{\mu}_2^*(0) = \underline{\mu}(T_2^*(\overline{\mu}_2^*(0)))$, imply that $\overline{\mu}_2^*(\tau) = \underline{\mu}(\tau)$, contradiction.

Next, consider the case where $k > \overline{k}^*$ and (4) holds, such that regime 3 applies for $\tau = 0$. Here, we distinguish two cases, depending on whether there are two payments 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 4, 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:[41]

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

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, from standard monotone comparative statics results, 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 4, $\frac{\partial T}{\partial \tau} < 0$, 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

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

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

Using the optimality condition for T (cf. 12) these expressions can be simplified to obtain

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

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. 12) these expressions can be rewritten to obtain

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

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

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 7. If $k \geq \underline{k}^*$, it holds from Proposition 6 that $\mu^*(\tau)$ is strictly increasing in τ and, thus, maximized at the highest possible value $\tau = \overline{\tau}$. If $k < \underline{k}^*$, we know from Proposition 5 that $\mu_1^*(\tau)$ is, for $\tau < \widetilde{\tau}$, decreasing and, for $\tau > \widetilde{\tau}$, increasing in τ. Hence, it must be maximized at either $\tau \in [0, T_1(\overline{\mu}_1^*)]$ 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 $\overline{\mu}_1^*(0) < \lim_{\tau\to\infty} \mu_1^*(\tau) = \overline{\mu}$. Now, from (13) it holds that $\overline{\mu}$ is strictly increasing in k and approaches zero for $k \to 0$, while $\overline{\mu}_1^*(0)$ is, from the slack acquisition constraint, independent of k. Hence, there exists a unique value $\widetilde{k} = c'(\overline{\mu}_1^*(0))\overline{\mu}_1^*(0) - c(\overline{\mu}_1^*(0))$, such that $\overline{\mu}_1^*(0) \geq \overline{\mu}$ for $k \leq \widetilde{k}$, and $\overline{\mu}_1^*(0) < \overline{\mu}$ 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 $\overline{\mu}_1^*(0)$ and $\overline{\mu}_1^*(\overline{\tau})$. Still, as, for $\overline{\tau} < \infty$, $\overline{\mu}_1^*(0) \geq \overline{\mu}$ implies $\overline{\mu}_1^*(0) > \overline{\mu}_1^*(\overline{\tau})$, while similarly $\overline{\mu}_1^*(0) \leq \overline{\mu}_1^*(\overline{\tau})$ implies $\overline{\mu}_1^*(0) < \overline{\mu}$, the cutoff result continues to hold. **Q.E.D.**

7 Appendix B: Additional Material

In this Appendix, we relate our analysis of the optimal compensation without regulation (and for fixed level of diligence) more closely to the literature. For this we first solve the firm's problem now with binary diligence effort $\mu \in \{\mu_l, \mu_h\}$, where we take as given that the firm wants to implement μ_h (as well as $a_h = 1$). Denoting the costs of diligence effort by $c(\mu_l) = c_l$ and $c(\mu_h) = c_h$, the firm's problem reads

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

$$s.t.$$

$$\sum\nolimits_i b_i e^{-r_A T_i} \left(e^{-(\lambda_H - \Delta_\lambda) T_i} - e^{-\lambda_H T_i} \right) \geq \frac{c_h - c_l}{\mu_h - \mu_l}, \tag{27}$$

$$\sum\nolimits_i b_i e^{-r_A T_i} \left[\mu_h e^{-(\lambda_H - \Delta_\lambda) T_i} + (1 - \mu_h) e^{-\lambda_H T_i} \right] - c_h \geq k, \tag{28}$$

$$T_i \geq 0,$$

$$b_i \geq 0.$$

This is essentially the same problem as in the case with continuous diligence effort, with the only difference that the incentive constraint for μ is no longer determined from the first-order approach, but given by the inequality in (27). However, it is easy to show that this constraint will always bind under the optimal contract implementing μ_h. Suppose to the contrary. Then, one could always reduce payments in $T_i > 0$ until (27) binds and increase the up-front payment in $T_0 = 0$ to still satisfy (28), which reduces compensation costs due to the wedge in discount rates Δ_r. This leads to a contradiction of optimality. Hence, the optimal compensation scheme with binary diligence effort is analogous to the one characterized in Proposition 1 for continuous diligence effort, once we set $\mu = \mu_h$, $c(\mu) = c_h$, and $c'(\mu) = \frac{c_h - c_l}{\mu_h - \mu_l}$.

Comparison with Hartman-Glaser et al. (2012). In their model the underwriting bank (as the agent) can exert discrete, unobservable effort that reduces the likelihood of default, where the default time follows an exponential distribution.[42] Their parameter restriction essentially implies $\mu_h = 1$ and $\mu_l = c_l = 0$. This specification implies crucially

[42] Their contribution is, however, wider in a different aspect. They show that when the agent (i.e., the bank in their model) makes more than one loan, than the optimal contracts rewards the agent only when none of these loans defaulted until time T.

that (4) can not hold, so that an up-front payment is never optimal. Further, in the Appendix to their paper, Hartman-Glaser et al. (2012) provide an informal extension allowing low effort (μ_l) to generate a convex combination of high and low default rates. This does not allow for $\mu_h < 1$, however, such that also with this parameterization (4) never holds.

Comparison with Malamud et al. (2013). Malamud et al. (2013) also consider optimal contracting in a setting where an intermediary can exert costly ex-ante effort to reduce the default risk of some financial asset, allowing for a large class of default time distributions.[43] As effort in their model can take on only finitely many values, we will now compare their results with the binary diligence effort version of our model outlined above. Further, while their model allows for the intermediary to be risk-averse, we restrict attention to the results they obtain in the limit as risk-aversion goes to zero.

One of their main results lies in defining an upper bound on the number of payments under the optimal contract depending on the default time distribution (cf. their Theorem 4.3). Applying their results to the setting of our model,[44] it is straightforward to show that this upper bound is finite, implying that our restriction to a grid of countably many payment dates is without loss of generality. (In particular, this rules out the optimality of paying rates.) What remains to be shown is in which cases or whether at all this upper bound is actually achieved under the optimal contract and when precisely payments occur. Malamud et al. (2013) provide such a complete characterization for the binary effort model with default time distributions from the Black and Cox (1976) model (cf. their Theorem 4.5), which, however, does not nest our specification.[45] Still, their findings are complementary to ours in that there are cases where two payments are optimal, one at time zero and another one at some $T > 0$, showing that some of the core features of our optimal compensation design also hold for a different class of default time distributions.

[43] A complete characterization of the optimal contract with a risk-neutral intermediary is derived for distributions from the Black and Cox (1976) model.

[44] Theorem 4.3. applies to our setting, since both players are risk-neutral and the mixed-exponential distribution implies that higher effort leads to a lower hazard rate.

[45] Further, as is easily verified, also their additional characterization results in Propositions 4.6 to 4.10 do not apply to our specification of the default time distribution.