Termination as an incentive device*

Borys Grochulski[†] Yuzhe Zhang[‡]

April 4, 2022

Abstract

In this paper, we study the conditions under which termination is a useful incentive device in the canonical dynamic principal-agent moral hazard model of Sannikov (2008). We find that temporary suspension of the agent after poor performance dominates termination if the principal's outside option is low and the agent's outside option is moderate. In suspension, the agent performs tasks free of moral hazard and receives no compensation, which rebuilds his "skin in the game" and allows for incentives to be restored without terminating. If the agent's outside option is low, suspension is ineffective because it rebuilds the agent's skin in the game too slowly. If the agent's outside option is high, the profitability of the relationship with the agent is low, so the principal prefers to terminate rather than extend the relationship through temporary suspension. Because the optimal use of suspension versus termination after poor performance can be highly sensitive to the principal's and agent's outside options, similar jobs can have vastly different average job durations, purely for incentive reasons.

Keywords: incentives, dynamic moral hazard, termination, suspension, slow reflection

JEL codes: D86, D82, M55, C61

1 Introduction

A clear lesson obtained in the literature on agency problems is that, in order to respond to incentives, the agent must maintain a stake in the relationship, or "skin in the game."

*The authors would like to thank the editor and three anonymous referees as well as Sebastian Di Tella, Huberto Ennis, Boyan Jovanovic, Tuomas Laiho, Christopher Phelan, David Rahman, Yuliy Sannikov, Martin Szydlowski, Juuso Välimäki, Jan Werner, Russell Wong, and John Zhu for their helpful comments, and Jackson Evert for excellent research assistance. An earlier version of this paper was circulated as "Optimal contracts with reflection." The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[†]Federal Reserve Bank of Richmond, borys.grochulski@rich.frb.org.

[‡]Texas A&M University, zhangeager@tamu.edu.

Incentives are provided by exposing the agent's stake to performance risk, where the agent's stake is increased if performance is strong and decreased if performance is weak. It is much less clear, however, what should be done when the agent's stake runs out, i.e., when a streak of poor performance reduces the agent's stake to his outside option, his participation constraint becomes binding, and his stake cannot be reduced any further. Many standard agency models, including the canonical dynamic principal-agent model of Sannikov (2008), assume that the relationship is at that point terminated.

In this paper, we show that this assumption is restrictive. In many cases, temporary suspension of the agent is preferable to terminating the relationship. The trade-off between terminating and suspending the agent after poor performance depends on the outside options of the principal/firm and the agent. Using the model of Sannikov (2008), we show that suspending the agent when his stake in the relationship runs out is preferable to terminating if the firm's outside option is low and the agent's outside option is moderate.

As defined in Zhu (2013), suspension is a contractual phase in which the agent receives no compensation and exerts no effort, so no incentives are required. The desired effect of suspension is a deterministic increase of the agent's stake in the relationship, which allows for incentives to be restored at the end of suspension without terminating. The agent's stake in the relationship is measured by his continuation value in the contract, and, more precisely, by the excess of his continuation value over his outside option. Suspension is feasible, i.e., achieves its desired effect of increasing the agent's stake, only if the agent's flow of utility in suspension is strictly lower than his continuation value in the contract. In particular, when the agent's stake in the relationship runs out, i.e., when his continuation value in the contract matches his outside option, suspension is feasible only if the agent's flow of utility in suspension is strictly lower than his outside option.

In our model, the agent's flow of utility when receiving no compensation and exerting no effort is normalized to zero. When the agent's stake in the relationship runs out, therefore, suspension is feasible only if the agent's outside option is strictly positive. If the agent's outside option is not positive, suspension does not restore the agent's skin in the game, and, thus, termination after poor performance is, trivially, optimal.

The agent's positive outside option is a necessary condition for the optimality of suspension, but it is not sufficient because both the benefit and the cost of using suspension decrease with the level of the agent's outside option. The benefit of using suspension rather than terminating is the profitability of the relationship, which is restored as soon as suspension ends. A high agent outside option reduces this benefit because the relationship is less profitable when the agent's participation constraint is tighter. The cost of using suspension rather than terminating is the zero expected flow of output the firm receives during suspension (implied by the agent's zero effort and zero compensation), which is lower than the flow value of the firm's outside option. A high agent outside option reduces this cost by reducing the expected duration of suspension. Indeed, suspension is shorter when the agent's flow of utility in suspension (normalized to zero) is farther below his continuation value (which during suspension is equal to his outside option).

If the agent's outside option is very high, its impact on the benefit of using suspension is strong: the agent's participation constraint becomes so tight that the firm cannot match the agent's outside option without turning in a loss. In this case, clearly, suspension has no value as the firm prefers to not offer a contract to the agent at all. If the agent's outside option is positive but low, its impact on the cost of suspension is strong. In particular, if the agent's outside option is zero, the flow of utility delivered to the agent in suspension exactly matches his outside option, the duration of suspension becomes infinite, and suspension becomes useless. Suspension, thus, is most valuable if the agent's outside option is moderate, i.e., neither too high nor too low.

Interestingly, since the benefit of a shorter suspension can outweigh the cost of a tighter participation constraint, an optimal contract may terminate an agent with a low outside option while suspending and retaining an otherwise identical agent whose outside option is higher. Thus, an increase in an agent's outside option can make it less likely that the agent will leave the firm.

In sum, an optimal contract will suspend the agent after poor performance instead of terminating if the agent's outside option is moderate and the firm's outside option is sufficiently low, as a lower firm outside option reduces the opportunity cost of suspending. In particular, if the firm's outside option is zero, suspension has no opportunity cost, which makes suspension preferable to termination in all cases, as long as suspension is feasible and the overall relationship is profitable.

With suspension after poor performance dominating termination, the optimal contract changes qualitatively: the contract's only exit is the standard retirement of the agent after strong performance, when the agent's continuation value reaches an endogenous retirement threshold.[1] As an implication, the model predicts that observed contract durations should be discretely longer when suspension is optimal. In fact, a small change in either the firm or the agent's outside option can alter the optimal choice of termination versus suspension after poor performance. In these cases, such a small change will trigger a large change in the expected duration of the contract. Thus, our standard moral hazard model shows that two similar jobs can have

[1] Clearly, our model isolates moral hazard. In reality, terminations after poor performance can be observed for reasons other than incentives. E.g., terminations can be due to learning of the agent's type or in response to a persistent negative shock to the quality of the match between the firm and the agent. Following Mortensen and Pissarides (1994), such shocks have been explored in models with long-term contracts in, e.g., Lamadon (2016) and Lise et al. (2016).

vastly different average job durations.

To relate the conditions for the optimal use of termination and suspension to external markets faced by the firm and the agent, it is easy to embed our optimal contracting problem in a simple search and matching framework, where the outside options in the current match are determined as the values the firm and the agent can obtain from a new match less search costs (see Remark 1 in Section 2). Our model predicts that termination of the agent after poor performance should be observed in jobs in which either the firm or the agent can rematch easily, while suspension should be observed in jobs in which the firm faces relatively high cost to replace the agent.

The model's predictions provide novel testable implications: suspensions should be observed in mid-level jobs, while terminations in low- and top-level jobs. Indeed, these implications follow if firms face relatively low costs to replace workers in low-level jobs, while workers in top-level jobs have high outside options.[2] Mid-level, skilled jobs fit the conditions for the optimality of suspension rather than termination if firms face substantial costs in finding suitable replacement and the workers' outside options are positive but moderate. Correspondingly, our model implies that higher average job durations should be observed in mid-level jobs than in either low- or top-level jobs.

In our model, suspension can be implemented only with the action of zero effort, which gives suspension a narrow meaning of doing nothing and receiving no compensation. More broadly, however, suspension can be implemented by a) switching the agent's job assignment to any task that is free of moral hazard and b) underpaying him.[3] Underpaying the agent in suspension allows his stake in the relationship, or continuation value, to increase. In fact, the more severely the agent is underpaid, the faster his stake in the relationship grows and, hence, the sooner he can exit suspension, which makes suspension more effective. Similarly, if multiple tasks free of moral hazard can be assigned to the agent, the more punishing the agent's assignment during suspension, the more effective suspension becomes. For example, if the firm could assign "busy work" to the agent (a task free of moral hazard that may be totally unproductive but is costly for the agent to perform), busy work would be assigned in suspension because such an assignment would reduce the necessary duration of suspension, thus making suspension more effective. Furthermore, if busy work is more tedious to the agent than regular work, the agent can be underpaid in suspension without a nominal decrease in compensation.

We characterize an optimal contract assuming commitment to no-renegotiation. When we relax

[2] Jenter and Lewellen (2020) argue that performance-induced CEO turnover is common. At the same time, following Jensen and Murphy (1990), the empirical literature on executive compensation, surveyed in Edmans and Gabaix (2016), finds that the threat of termination is not a very significant source of managerial incentives, consistent with high outside options of CEOs.

[3] For example, a professor may be assigned a high teaching load after a period of low output in research.

this assumption, the optimal contact can take two forms. If the firm's outside option is positive, the standard renegotiation-proof contract of DeMarzo and Sannikov (2006) with stochastic termination at an endogenous lower bound is optimal. If the firm's outside option is zero, for example due to vacancy creation costs in a competitive labor market, an optimal renegotiation-proof contract suspends the agent at an endogenous lower bound without terminating.

Relation to the literature In two-period principal-agent moral hazard models, the optimality of endogenous contract termination after poor performance is studied in Stiglitz and Weiss (1983) and Spear and Wang (2005). These two-period settings, however, are too simple to capture the benefits of a temporary suspension and a subsequent resumption of effort, which is an important alternative to termination in our model.

Reflection of the agent's continuation value process off a lower bound implied by the agent's participation constraint is optimal in many dynamic risk-sharing problems with private information, as shown in, e.g., Atkeson and Lucas (1995), Phelan (1995), Wang (1995), Fong and Li (2017), and Zilberman et al. (2019). These studies, however, do not consider terminations. Other studies, by contrast, make assumptions that eliminate suspension. For example, Clementi and Hopenhayn (2006) assumes that the agent's outside option is zero. Szydlowski (2019) assumes that the firm's outside option is sufficiently high. DeMarzo and Sannikov (2006) and He (2009) assume that the agent's shirking action is extremely costly for the firm. We allow for both suspension and termination and study the conditions under which one dominates the other.

In the model of DeMarzo and Sannikov (2006), Zhu (2013) relaxes the assumption that the agent's shirking action is extremely costly to the principal. He considers a class of models parametrized by how costly the shirking action is to the principal and how much utility it provides to the agent, and studies the conditions under which asking the agent to shirk is optimal. He finds that occasional shirking is a part of an optimal contract if the agent's shirking action is not too costly to the principal and gives either a sufficiently high or a sufficiently low flow of utility to the agent. Suspension of the agent occurs in an optimal contract in the latter case. In this paper, we follow Zhu (2013) by using two HJB component equations to construct an optimal contract: one equation restricted to maintain positive volatility in the agent's continuation value process, and one restricted to keep this volatility at zero. In our model, however, shirking always provides a zero flow payoff to the principal and a zero flow of utility to the agent. Under this assumption on the shirking action, we investigate how the trade-off between termination and suspension depends on the outside options of the principal and the agent. We find that the agent's outside option has a nonmonotonic impact on the value of suspension. The optimal timing of suspension is different in our model from that in Zhu (2013). In our model, it is never optimal to suspend the agent before his participation constraint binds. In Zhu (2013), the principal is more patient than the agent. Since suspending the agent entails an immediate cost and a delayed benefit, a patient principal is willing to use suspension sooner, i.e., before the agent's participation constraint binds.[4]

Piskorski and Westerfield (2016) introduce a costly stochastic monitoring technology to the model of DeMarzo and Sannikov (2006) and study how incentives are optimally provided by a mix of monitoring and standard pay-for-performance. Monitoring can detect shirking, and the agent faces a stigma (a continuation value below his outside option) if his shirking is detected. In the optimal contract, termination does not occur if monitoring is inexpensive, the stigma is sufficiently high, and the principal's outside option is sufficiently low. The alternative to termination, however, is not suspension but rather strong monitoring of the agent, where the probability of detection of shirking is high enough to allow for pay-for-performance incentives to be reduced to zero. Termination is still necessary in Piskorski and Westerfield (2016) if monitoring is sufficiently costly or the stigma attached to detected shirking is low. In our model, despite monitoring being very costly (i.e., not possible) and no stigma, termination is not necessary because the agent can be suspended. Suspension does not require monitoring because incentives are withdrawn in suspension. In Piskorski and Westerfield (2016), incentives remain switched on at all times and, thus, strong monitoring is necessary in any contract that does not terminate after poor performance.

Termination after poor performance is necessary in many optimal contracting environments with additional frictions. This is the case, e.g., in MacLeod and Malcomson (1989), where the agent's performance is not contractible, in Levin (2003), Fuchs (2007), and Zhu (2018), where the agent's performance can only be subjectively evaluated by the principal, or in Halac (2012), where the principal's outside options are private information.

Organization Section 2 lays out the model. Section 3 considers a baseline case in which the firm's outside option is zero. Section 4 considers the general case. Section 5 considers renegotiation-proof contracts. Section 6 concludes. The Appendix contains the proofs.

2 A dynamic principal-agent problem with moral hazard

Consider the canonical dynamic moral hazard principal-agent problem formulated in Sannikov (2008). A principal/firm owns a project that can produce a stream of output if operated by an agent. The firm can hire an agent who has an outside option with value B. While under contract to operate the project, the agent takes private actions that influence the project's

[4] In particular, the super-contact condition between the two component HJB solutions does not hold in our model. In Zhu (2013), the super-contact condition typically holds at an interior point, above the agent's outside option, which determines the point of suspension in his model.

output. In particular, cumulative output produced up to date t, X_t, follows

$$dX_t = A_t^a dt + \sigma dZ_t,$$

where $A_t^a \in \mathcal{A}$ is the agent's action (effort), Z_t is a standard Brownian motion on $(\Omega, \mathcal{F}, P)$, and $\sigma > 0$ is a constant. We assume that the set of feasible actions $\mathcal{A}$ is a compact interval $[0, \bar{A}]$ for some $\bar{A} > 0$.

The firm's outside option is its project's residual value $R \geq 0$, which the firm collects when the agent is terminated (i.e., fired) or retired.[5] If the agent is terminated, he collects his outside option B. Let τ_{tn} be a stopping time denoting the time of the agent's termination. If the agent is retired, the firm delivers to him some retirement value $W_{gp} \geq B$.[6] Let τ_{gp} be a stopping time denoting the time of the agent's retirement. We assume a retired agent does not collect the outside option value B.[7]

Remark 1 *One possible interpretation of B and R comes from a simple search and matching environment similar to Pissarides (1985) and Mortensen and Pissarides (1994). In that environment, B represents the agent's value of unemployment and job search, and R represents the firm's value of searching for a new agent. Let W_0 denote the agent's value at the onset of a contract and $V(W_0)$ the corresponding profit for the firm. In reduced form, search frictions can be thought of as driving a wedge between B and W_0 for the agent, and between R and $V(W_0)$ for the firm. Let the agent's and the firm's search costs be denoted by, respectively, κ_a and κ_f. Then, the outside option values B and R are determined, jointly with W_0 and the value function V, as a solution to*

$$B = (1 - \kappa_a)W_0 \quad \text{and} \quad R = (1 - \kappa_f)V(W_0). \tag{1}$$

A contract specifies stopping times τ_{tn} and τ_{gp}, the agent's retirement value W_{gp}, and a pair of progressively measurable processes $\{(C_t, A_t); 0 \leq t < \min\{\tau_{tn}, \tau_{gp}\}\}$, where $A_t \in \mathcal{A}$ is the action recommended for the agent to take at t, and $C_t \geq 0$ is his compensation. Compensation C_t cannot be negative due to the agent's limited liability.

[5] The residual value R could be coming from liquidation or from replacement of the agent with a new one after incurring some search costs, as in Remark 1. We assume the firm has the option to not hire an agent and has free disposal of the project, i.e., R is nonnegative.

[6] It is without loss of generality to restrict attention to contracts in which W_{gp} is constant. If we allow W_{gp} to be an arbitrary adapted process, we can show, similar to Sannikov (2008), that it is optimal to promise the same W_{gp} in all histories in which the agent is retired.

[7] Our results go through under the alternative assumption. In particular, condition (21) continues to capture the trade-off between termination and suspension at B. In the framework of Remark 1, our assumption means a retired agent does not rejoin the labor market, i.e., does not search for a new match. This assumption is consistent with models of labor market equilibrium, e.g., Wang (2011), where, due to moral hazard, total surplus in a match is zero if the agent's wealth, accumulated in his past matches, is sufficiently large.

At each $t < \min\{\tau_{tn}, \tau_{gp}\}$, the agent chooses privately his action $A_t^a \in \mathcal{A}$ to maximize his utility. A contract is incentive compatible if $A_t^a = A_t$ at all t, i.e., the actual action chosen by the agent is that recommended by the contract.

The agent's expected value from an incentive compatible contract is

$$\mathbb{E}\left[r \int_0^{\min\{\tau_{tn},\tau_{gp}\}} e^{-rt} \left(u(C_t) - h(A_t)\right) dt + 1_{\{\tau_{tn}<\tau_{gp}\}} e^{-r\tau_{tn}} B + 1_{\{\tau_{gp}<\tau_{tn}\}} e^{-r\tau_{gp}} W_{gp} \right],$$

where $r > 0$. The agent's utility function $u : \mathbb{R}_+ \to \mathbb{R}_+$ has a continuous second derivative with $u' > 0$, $u'' < 0$, $u(0) = 0$, $\lim_{c\to\infty} u(c) = \infty$, $\lim_{c\to 0} u'(c) = \infty$, and $\lim_{c\to\infty} u'(c) = 0$. The function $h : \mathcal{A} \to \mathbb{R}_+$ represents the agent's disutility from effort. We assume that its second derivative h'' is continuous, $h' > 0$, $h'' > 0$, $h(0) = 0$, and $\lim_{a\to 0} h'(a) =: \gamma_0 > 0$.

More generally, for any $t < \min\{\tau_{tn}, \tau_{gp}\}$, an incentive compatible contract defines the agent's continuation value process as

$$W_t \equiv \mathbb{E}_t \left[r \int_t^{\min\{\tau_{tn},\tau_{gp}\}} e^{-r(s-t)} \left(u(C_s) - h(A_s)\right) ds + 1_{\{\tau_{tn}<\tau_{gp}\}} e^{-r(\tau_{tn}-t)} B + 1_{\{\tau_{gp}<\tau_{tn}\}} e^{-r(\tau_{gp}-t)} W_{gp} \right].$$

A contract satisfies the agent's participation constraint if

$$W_t \geq B \quad \text{at all } t \geq 0. \tag{2}$$

The retirement value $W_{gp} \geq 0$ is (optimally) delivered to the agent by constant compensation, $c_{gp} \geq 0$, at all $t \geq \tau_{gp}$, where c_{gp} satisfies

$$W_{gp} = r \int_0^\infty e^{-rt} u(c_{gp}) dt = u(c_{gp}).$$

The firm's profit from delivering to the agent the retirement value $W_{gp} \geq 0$, therefore, is $R + F_0(W_{gp})$, where

$$F_0(W_{gp}) \equiv -u^{-1}(W_{gp}) = -c_{gp} \leq 0. \tag{3}$$

The firm's ex ante expected profit from an incentive compatible contract is

$$\mathbb{E}\left[r \int_0^{\min\{\tau_{tn},\tau_{gp}\}} e^{-rt} (A_t - C_t) dt + 1_{\{\tau_{tn}<\tau_{gp}\}} e^{-r\tau_{tn}} R + 1_{\{\tau_{gp}<\tau_{tn}\}} e^{-r\tau_{gp}} \left(R + F_0(W_{gp})\right) \right].$$

An optimal contract maximizes the firm's ex ante expected profit subject to incentive compatibility and the agent's participation constraint (2). Note that a degenerate contract with $W_0 = B$ and $\tau_{tn} = 0$ satisfies these constraints, i.e., the firm has the option to not hire the agent and collect the project's residual value R immediately at $t = 0$.

2.1 Recursive formulation of the contracting problem

Outside of Section 5, where we discuss renegotiation-proof contracts with stochastic termination, it is without loss of generality in our model to exclude jumps in the agent's continuation value at termination or retirement. Thus, we restrict attention to termination policies such that $\tau_{tn} < \infty$ implies $W_{\tau_{tn}} = B$ and retirement policies such that $\tau_{gp} < \infty$ implies $W_{\tau_{gp}} = W_{gp}$.

Following Sannikov (2008), we will use the agent's continuation value W_t as the state variable with a diffusion representation at all $t < \min\{\tau_{tn}, \tau_{gp}\}$:

$$dW_t = r(W_t - u(C_t) + h(A_t))dt + rY_t(dX_t - A_t dt), \tag{4}$$

where $dX_t - A_t dt$ is the agent's performance relative to the benchmark $A_t dt$, and $\{Y_t; 0 \le t < \min\{\tau_{tn}, \tau_{gp}\}\}$ is a progressively measurable performance-sensitivity process.[8] It is a standard result that a contract is incentive compatible if

$$A_t \in \underset{a \in \mathcal{A}}{\operatorname{argmax}} \{Y_t a - h(a)\} \tag{5}$$

at all $t < \min\{\tau_{tn}, \tau_{gp}\}$. Note that this condition implies that action $A_t = 0$ is incentive compatible if and only if $Y_t \le \gamma_0$, $A_t \in (0, \bar{A})$ is incentive compatible if and only if $Y_t = h'(A_t) > \gamma_0$, and $A_t = \bar{A}$ is incentive compatible if and only if $Y_t \ge h'(\bar{A})$.

Using the state variable W_t, Sannikov (2008) has expressed termination and retirement policies as first passage times of W_t to, respectively, B and W_{gp}. While $\tau_{gp} = \min\{t : W_t = W_{gp}\}$ is without loss of generality, $\tau_{tn} = \min\{t : W_t = B\}$ is. We will verify that for a generic set of values of B and R, $\tau_{tn} = \infty$ is optimal, i.e., the agent is never terminated.

For given τ_{tn} and τ_{gp}, standard dynamic programming arguments imply the existence of a concave value function $V : [B, \infty) \to \mathbb{R}$, which represents the firm's continuation profit under an optimal contract at all $t < \min\{\tau_{tn}, \tau_{gp}\}$. This V satisfies the following HJB equation:

$$V(W_t) = \max_{c \ge 0, a \in \mathcal{A}, Y} \{a - c + V'(W_t)(W_t - u(c) + h(a)) + \frac{1}{2}V''(W_t) r\sigma^2 Y^2\}, \tag{6}$$

where controls a and Y jointly satisfy the incentive compatibility constraint (5).

To find an optimal contract, we follow a standard three-step approach. First, we solve the HJB equation (6) to obtain a candidate, v, for the value function V. In doing so, we impose

[8] In Section 5, we allow for a jump in W_t at termination and modify the representation (4) accordingly.

appropriate boundary conditions at the agent's participation and retirement thresholds. In particular, since terminating the agent is an option when $W = B$, $v(B) \geq R$. Since retiring the agent is always an option, $v(W) \geq R + F_0(W)$ for all W. Second, from the candidate solution v, we construct a contract using the policy functions $c(\cdot)$, $a(\cdot)$, and $Y(\cdot)$ that attain the solution v in the HJB equation (6). In particular, the agent's continuation value process associated with the candidate solution v is given as a (weak) solution to the stochastic differential equation

$$dW_t = r\left(W_t - u(c(W_t)) + h(a(W_t))\right)dt + rY(W_t)\left(dX_t - a(W_t)dt\right) \tag{7}$$

for $0 \leq t < \min\{\tau_{tn}, \tau_{gp}\}$, with some fixed initial value $W_0 \in [B, W_{gp}]$. Third, we verify that this contract is optimal, i.e., $v(W_0) = V(W_0)$ for all $W_0 \in [B, W_{gp}]$.

2.2 High- and low-action ODEs

Following the approaches of Sannikov (2008) and Zhu (2013), it will be useful for us to write the HJB equation (6) as follows

$$V(W) = \max\Big\{ \max_{c \geq 0, a \in \mathcal{A}} \{a - c + V'(W)(W - u(c) + h(a)) + \frac{1}{2}V''(W)r\sigma^2(h'(a))^2\}, \tag{8}$$

$$\max_{c \geq 0}\{-c + V'(W)(W - u(c))\}\Big\}. \tag{9}$$

Here, we are writing out separately the option of using high volatility $Y \geq \gamma_0$, in line (8), and the option of using low volatility $Y \in [0, \gamma_0)$, in line (9). The outside maximization over these two options makes this formulation equivalent to (6). Furthermore, note that in (8) we have used the incentive compatibility constraint (5) to substitute Y with $h'(a) \geq \gamma_0$. Similarly, in (9) we have used the fact that incentive compatibility requires $a = 0$ when $Y < \gamma_0$, and, further, that, with a concave V, zero volatility $Y = 0$ dominates any volatility $Y \in (0, \gamma_0)$.

Following Zhu (2013), we will study the two options in (8) and (9) as two separate ordinary differential equations (ODEs), whose solutions, denoted respectively as F and L, will be combined to derive an optimal contract:

$$F(W) = \max_{c \geq 0, a \in \mathcal{A}} \{a - c + F'(W)(W - u(c) + h(a)) + \frac{1}{2}F''(W)r\sigma^2(h'(a))^2\}, \tag{10}$$

$$L(W) = \max_{c \geq 0}\{-c + L'(W)(W - u(c))\}. \tag{11}$$

The first ODE, (10), is exactly the equation studied in Sannikov (2008). Contracts derived from solutions to this equation, due to the restriction $Y = h'(a) \geq \gamma_0 > 0$, have strictly positive volatility of W_t at all $t < \min\{\tau_{tn}, \tau_{gp}\}$.[9] Although it could be natural to call (10) positive-volatility ODE, in order to emphasize the analogy with Zhu (2013), we will refer to (10) as the high-action ODE.

The second ODE, (11), forces the volatility Y to be zero and uses the no-effort action $a = 0$ at all times. We will call this ODE low-action ODE. The advantage of having $Y = 0$ is that along any solution to the low-action ODE, the dynamics of W_t are deterministic, i.e., W_t is not sensitive to output. This property is used in Section 3.2, where we construct the suspension phase of the contract.

3 Termination versus suspension with $R = 0$

To examine the optimal use of termination and suspension after poor performance, we start with a special case in which the project has no residual value, and, thus, the firm's outside option is zero. This case is important for two reasons. First, it is consistent with equilibrium in models with search frictions, e.g., Pissarides (1985) and Mortensen and Pissarides (1994), where vacancy creation costs and competition from other firms reduce to zero the ex ante value of creating a new vacancy. Second, it is instructive to examine the optimal provision of incentives via termination or suspension in isolation from the firm's concern for delaying the receipt of $R > 0$, which we examine in the general case in Section 4.

In this section, thus, we fix $R = 0$ and allow the agent's outside option B to be any number, including negative.[10] Clearly, if the agent's outside option B is sufficiently high, the relationship will not form because the productivity of the project is not sufficient to meet the agent's outside option and turn in a profit. We therefore restrict attention to $B < \hat{B}$, where $\hat{B}$ is a loose upper bound defined in Appendix A.1.

In Section 3.1, we follow Sannikov (2008) in assuming termination at B. In Section 3.2, we show that this assumption is restrictive: in every profitable relationship, suspension of the agent after poor performance dominates termination so long as suspension is feasible. In Section 3.3, we characterize the level of premium afforded by suspending the agent at B rather than terminating. Section 3.4 gives a formal statement and verification of the optimal contract.

[9] Note that action $a = 0$ is allowed in (10) but only with positive volatility $Y = h'(0) = \gamma_0 > 0$. The pair $(a, Y) = (0, \gamma_0)$, although incentive compatible, is never used in the optimal contract. Instead, Lemma A.3 in the Appendix implies that action $a = 0$ is used only with volatility $Y = 0$, when the optimal contract is determined by a solution to the other component ODE, (11). The pair $(a, Y) = (0, \gamma_0)$ is allowed in (10) merely for technical reasons.

[10] Negative B can arise if being fired carries a stigma (negative utility). Also, if minimum wage laws impose a strictly positive lower bound on compensation, being fired can make the agent worse off than earning the minimum wage and providing zero effort.

Section 3.5 discusses the dynamics of the reflection of the agent's continuation process at B under the optimal contract.

3.1 Optimal contract with termination

In this subsection, we follow Sannikov (2008) in assuming termination after poor performance, i.e., when the agent's continuation value has reached B. The optimal contract with termination at B is constructed from a solution F to high-action ODE (10) obtained with two specific boundary conditions. The first boundary condition is a value-matching condition at B. Since the contract terminates at B and the residual value of the project is zero, we have

$$F(B) = 0. \tag{12}$$

The second condition is a free-boundary condition for retirement of the agent. It requires that $F(W) \geq F_0(W)$ for all $W \geq B$, and

$$F(W_{gp}) = F_0(W_{gp}) \quad \text{and} \quad F'(W_{gp}) = F_0'(W_{gp}) \tag{13}$$

for some $W_{gp} \geq B$. Let us denote the solution to the high-action ODE (10) with boundary conditions (12) and (13) by $\tilde{F}$ and refer to it as the firm's profit function with termination at B.

Lemma 1 *For each $B < \hat{B}$, there exists a unique solution $\tilde{F}$ to the high-action ODE (10) with boundary conditions (12) and (13). The initial slope of $\tilde{F}$, $\tilde{F}'(B)$, is a continuous function of B. There exists a unique $\bar{\boldsymbol{B}} \in (0, \hat{B})$ such that*

$$\tilde{F}'(B) \begin{cases} > 0 & \text{for } B < \bar{\boldsymbol{B}}, \\ = 0 & \text{for } B = \bar{\boldsymbol{B}}, \\ < 0 & \text{for } B > \bar{\boldsymbol{B}}. \end{cases}$$

Proof The first two statements follow by setting $R = 0$ in Lemma 2. The last statement follows by setting $R = 0$ in the first part of Proposition 2. ■

For $B \geq \bar{\boldsymbol{B}}$, the termination profit function $\tilde{F}$ is decreasing, i.e., the boundary value $\tilde{F}(B) = 0$ is a global maximum of $\tilde{F}$. In these cases, despite $R = 0$, the firm does not offer a contract to the agent because the agent is too expensive to hire, i.e., the project the firm owns is not productive enough to deliver at least B to the agent and a positive profit to the firm.[11]

For $B < \bar{\boldsymbol{B}}$, the termination profit function $\tilde{F}$ is first increasing then decreasing, i.e., it

[11] We have $\bar{\boldsymbol{B}} < \hat{B}$ because the loose upper bound $\hat{B}$ derived in Appendix A.1 assumes no moral hazard.

has an interior peak. The optimal contract with termination is constructed from the policy functions c, a, Y that attain $\tilde{F}$ in the high-action ODE (10). The agent's promised value process starts at the peak of $\tilde{F}$, i.e., $W_0 = \text{argmax}_{W \geq B} \tilde{F}(W) > B$, and evolves according to (7). In particular, the agent's effort $A_t = a(W_t)$ is strictly positive and, correspondingly, the volatility of the agent's continuation value is strictly positive (bounded below by $\gamma_0 > 0$) at all $t < \min\{\tau_{tn}, \tau_{gp}\}$. The support of the state variable W_t is $[B, W_{gp}]$. When W_t hits either end of the interval $[B, W_{gp}]$, the contract ends, and the firm collects the residual value, R, which in this section is assumed to be zero. The agent is fired or retired, depending on which end of the interval was reached. Under this contract, thus, we have $\tau_{tn} = \min\{t : W_t = B\}$ and $\tau_{gp} = \min\{t : W_t = W_{gp}\}$.

Relative to the firm's first-best profit levels, which are attainable in the absence of moral hazard, termination of the relationship when $W_t = B$ is more costly than it is when $W_t = W_{gp}$, because at B the agent is owed little.[12] Therefore, it is intuitive that the firm should want to avoid terminating at B and should prefer a contract that preserves the relationship at that point. Indeed, we show in the next section that a fully optimal contract never terminates at B so long as a temporary suspension can be used to lift the agent's continuation value above B.

3.2 Optimal contract with suspension

We now relax the assumption that the agent is terminated at B. We discuss how, for any $B \in (0, \bar{\boldsymbol{B}}]$, a low-action ODE solution can be used to obtain a boundary condition for the high-action ODE solution that dominates the terminating boundary condition (12). As in Zhu (2013), the contract obtained from this superior boundary condition suspends the agent temporarily at B, without terminating. In suspension, the agent is asked for no effort and given no compensation. Our discussion here is informal with the objective of providing a guess for the optimal contract, which we formally verify in Theorem 1.

It is easy to verify that for any constant $\alpha \geq 0$, the ray out of the origin,

$$L(W) = \alpha W, \tag{14}$$

is a solution to the low-action ODE (11).[13] Because the policies that achieve L are $a = Y = c = 0$, we refer to the contract that implements L as suspension: the agent is not asked to work and is not paid any compensation. Substituting $a = Y = c = 0$ into (7) shows that the

[12] It is easy to verify that the gap between the firm's first-best profit function and F_0 is decreasing in W.

[13] Indeed, with $L'(W) = \alpha \geq 0$, the maximum in (11) is attained by $c = 0$. Thus, the right side of (11) reduces to αW.

agent's continuation value in suspension satisfies

$$dW_t = rW_t dt. \tag{15}$$

With $B > 0$, constraint (2) implies $W_t > 0$ at all t. Thus, the dynamics of W_t in suspension are very simple: W_t grows exponentially at the rate r. Intuitively, in the absence of volatility, the firm's obligation toward the agent is akin to a bond with the required rate of return of r. Since no payments are made to the agent in suspension, the balance owed to the agent must be increased at the rate r.

Next, we show how a low-action ODE solution in (14) can be used to construct an optimal contract, in which the agent is suspended rather than terminated at the lower bound $B \in (0, \bar{\boldsymbol{B}}]$. Figure 1 depicts the solution curve $\tilde{F}$ representing the firm's profit function assuming termination of the agent at B, and a low-action ODE solution, labeled as $\tilde{L}$, that is tangent to $\tilde{F}$ at some $W^s > B$. Since their levels and slopes are the same at W^s, the two solutions paste smoothly at that point. Consider now a contract (C, A) constructed by using the optimal controls from the low-action ODE solution $\tilde{L}$ at all $W \in [B, W^s]$ and the optimal controls from the high-action ODE solution $\tilde{F}$ at all $W \in (W^s, W_{gp}]$.[14] This contract delivers to the firm profit $\tilde{L}(W)$ if $W \in [B, W^s]$ and $\tilde{F}(W)$ if $W \in (W^s, W_{gp}]$. Because $\tilde{L}(W) > \tilde{F}(W)$ for all $W \in [B, W^s)$, the new contract constitutes a Pareto improvement over the optimal contract that terminates at B.

By (15), the process W_t implied by this contract is deterministic in the interval $[B, W^s]$. If initiated at some $W_0 < W^s$, the agent's continuation value W_t grows exponentially until it moves out of $[B, W^s]$. Once W_t leaves $[B, W^s)$, it never drops below W^s again because it grows deterministically, as in (15), in every future visit to W^s. This generates reflection off W^s whenever W_t reaches W^s from above.

Note also in Figure 1 that the second derivatives of $\tilde{L}$ and $\tilde{F}$ are not equal at W^s.[15] With $W^s > B$, the contract obtained by splicing $\tilde{L}$ and $\tilde{F}$ at W^s is not optimal, as better combinations of low- and high-action ODE solutions can be obtained by moving the splicing point closer to B. To see this, consider the high-action ODE solution F_1 depicted Figure 1. This solution strictly dominates the optimal solution with termination, $\tilde{F}$.[16] Note that $\tilde{L}''(W^s) > \tilde{F}''(W^s)$ implies that $\tilde{L}$ lies above $\tilde{F}$. With F_1 close enough to $\tilde{F}$, thus, $\tilde{L}$ crosses F_1 twice. Suppose now that we increase the slope of the low-action ODE solution until the two crossing points collapse to a single point, denoted in Figure 1 by W_1^s, at which solutions L_1 and F_1 are tangent. This pair of ODE solutions strictly dominates the pair $\tilde{L}$ and $\tilde{F}$, with the splicing point W_1^s located

[14] The stopping times associated with this contract are $\tau_{tn} = \infty$ and $\tau_{gp} = \min\{t : W_t = W_{gp}\}$.

[15] That is, $\tilde{L}$ and $\tilde{F}$ paste smoothly at W^s but violate the so-called super-contact condition.

[16] The point W_{gp} associated with F_1, not pictured, also exceeds the W_{gp} associated with $\tilde{F}$.

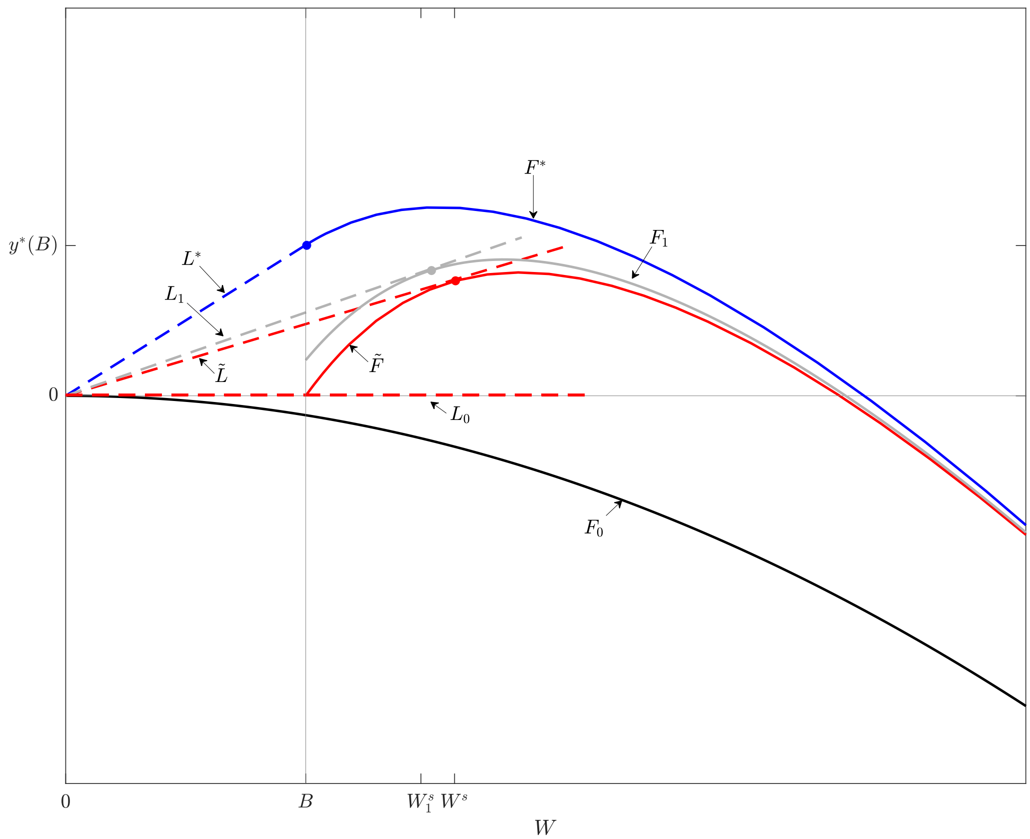

Figure 1: Suspension dominates termination at $B \in (0, \bar{\boldsymbol{B}}]$. The high-action ODE solution $\tilde{F}$ represents the optimal contract with termination at B. The low-action ODE solution L_0 is flatter than $\tilde{F}$ at $(B, 0)$ showing that an improvement on $\tilde{F}$ is possible. The low-action solution $\tilde{L}$ improves on $\tilde{F}$ by splicing at W^s. Solutions L_1 and F_1 move the splicing point closer to B. The optimal contract is represented by L^* spliced with F^* at $(B, y^*(B))$. In this example, $u(c) = \sqrt{c}, h(a) = 0.5a^2 + 0.4a$, $r = 0.1$, $\sigma = 1$, and $B = 0.1$.

closer to B than W^s.[17] [18]

Intuitively, a lower splicing point W^s allows the firm to attain a higher profit curve because it enlarges the (endogenous) support $[W^s, W_{gp}]$ for the state variable W_t in an optimal contract, which allows for the costly suspension or retirement of the agent to be delayed. Indeed, at W^s the contract must ask for zero effort, and at W_{gp} the project ends. With more distance between W^s and W_{gp}, the contract can sustain positive effort for longer and/or ask for higher

[17] Indeed, if $W_1^s > W^s$, then F_1 exceeds $\tilde{F}$ at W_1^s in both the level and the slope. Lemma A.1 in Appendix A.2 then implies that $F_1(W) > \tilde{F}(W)$ for all $W \geq W_1^s$ making it impossible for F_1 to reach F_0, thus violating the smooth-pasting condition (13).

[18] Note that if L is more concave than F, lowering the splicing point may not be optimal, and, thus, the optimal splicing point may be interior. Such cases are found in Zhu (2013), where the principal is more patient than the agent and the low-action ODE has strictly concave solutions.

levels of effort because the volatility of W_t necessary to induce high effort does not cause W_t to hit W^s or W_{gp} as quickly. In other words, a lower splicing point allows suspension to be postponed, which makes suspension less costly to implement ex ante.

Consistent with this intuition, the optimal contract is obtained when the splicing point W^s is set as low as possible, i.e., when W^s coincides with the lower bound B.[19] In this case, the splicing point cannot be moved further to the left, i.e., the endogenous support $[W^s, W_{gp}]$ cannot be made any larger or, equivalently, suspension cannot be delayed any further. In Figure 1, this solution is denoted by L^* spliced with F^* at the point $(B, y^*(B))$.

In general, it is optimal to maximize the curvature of the solution to the overall HJB equation (8)-(9). The low-action option L is chosen if it is more concave than F or at the boundary, where an application of F would violate the agent's participation constraint (2). The former case arises in Zhu (2013) and the latter in our model.

3.3 Finding the maximum boundary premium y^*

In the previous section, we showed that an optimal contract with suspension is generated by splicing a low-action ODE solution and a high-action ODE solution at $W^s = B$. In this section, we describe a procedure for finding the profit level at which the two ODE solutions are spliced, $y^*(B)$.

Let us fix $B \in (0, \bar{\boldsymbol{B}}]$ and take some $y \geq 0$. The solution $L = \alpha W$ to the low-action ODE that goes through the point (B, y) has slope $\alpha = y/B$. We look for y such that the high-action ODE solution F that splices smoothly with L at the point (B, y) also satisfies the optimal retirement condition (13). The smooth splicing conditions at the point (B, y) are

$$F(B) = L(B) = y \quad \text{and} \quad F'(B) = L'(B) = \frac{y}{B}. \tag{16}$$

We search for the splicing level y as follows. For each $y \geq 0$, the initial condition $F(B) = y$ and the optimal retirement condition (13) pin down a unique solution F to the high-action ODE by the forward-shooting argument of Sannikov (2008). Let us denote the initial slope of this solution, $F'(B)$, by $x(B, y)$.[20] We look for y at which the second condition in (16) is met as well, i.e., such y that $x(B, y) = y/B$.

Proposition 1 ***(Positive boundary premium from suspension)*** *For each $B \in (0, \bar{\boldsymbol{B}}]$,*

[19] Splicing points $W^s < B$ are inconsistent with the agent's participation constraint (2).

[20] Note that $x(B, 0) = \tilde{F}'(B)$.

$x(B, y)$ is continuous and strictly decreasing in y, and there exists a unique $y^* \geq 0$ such that

$$x(B, y^*) = \frac{y^*}{B}.$$

If $B = \bar{\boldsymbol{B}}$, then $y^* = 0$. For $B \in (0, \bar{\boldsymbol{B}})$, $y^* > 0$.

Proof Follows by setting $R = 0$ in Proposition 2. ■

If $y = 0$, then the low-action solution L through the point (B, y) is flatter at that point than the high-action solution $F = \tilde{F}$ because the slope of this L is zero and Lemma 1 implies $x(B, 0) \geq 0$ for all $B \leq \bar{\boldsymbol{B}}$, with a strict inequality for all $B < \bar{\boldsymbol{B}}$.[21] As we increase y, the low-action solution L through the point (B, y) becomes steeper, i.e., y/B increases, and the optimal retirement condition (13) forces the high-action solution F through the point (B, y) to become flatter, i.e., $x(B, y)$ decreases. If $y = Bx(B, 0)$, then the low-action solution L through the point (B, y) is steeper at (B, y) than the high-action solution F satisfying the optimal retirement condition (13).[22] By continuity and monotonicity, the slopes of L and F are equal at (B, y^*) for some unique $y^* < Bx(B, 0)$. Note that the case $B = \bar{\boldsymbol{B}}$ is special in that $x(\bar{\boldsymbol{B}}, 0) = 0$, i.e., the two slopes are equal already at the point $(\bar{\boldsymbol{B}}, 0)$ and, thus, $y^* = 0$ in this special case.

We will denote this unique y^* by $y^*(B)$ and the two solutions spliced at $(B, y^*(B))$ by, respectively, L^* and F^*. By $W_{gp}(B)$, we will denote the optimal agent retirement threshold W_{gp} pinned down by the solution F^*. For each $B \in (0, \bar{\boldsymbol{B}}]$, we have $F(B) = y^*(B) \geq 0$. Since termination at B yields $\tilde{F}(B) = 0$, we refer to $y^*(B)$ as the boundary premium that suspension generates over termination.

Proposition 1 does not apply to $B \leq 0$ or $B > \bar{\boldsymbol{B}}$. If $B \leq 0$, suspending the agent at B would violate the agent's participation constraint (2), as (15) implies that the agent's continuation value W_t would move downward from B.[23] At all $B \leq 0$, thus, suspension is infeasible, which makes termination trivially optimal.[24] If $B > \bar{\boldsymbol{B}}$, suspension moves the agent's continuation value upward, but doing so is not useful because hiring (or continuing a relationship with) an agent whose continuation value exceeds $\bar{\boldsymbol{B}}$ is not profitable for the firm. In sum, in any profitable relationship, suspension dominates termination so long as suspension is feasible.

[21] In Figure 1, this low-action solution is denoted by L_0.

[22] Indeed, if $y = Bx(B, 0)$, then $y/B = Bx(B, 0)/B = x(B, 0) > x(B, y)$, where the inequality follows from the strict monotonicity of x in y.

[23] Strictly so if $B < 0$. If $B = 0$, W_t would stay at B forever.

[24] As discussed in Section 3.2, suspending the agent before W_t reaches B is never optimal in our model. By contrast, such early suspensions are optimal in Zhu (2013), where the principal is more patient than the agent.

3.4 Optimal contract: formal statement and verification

Following Zhu (2013), we will define a function $v : [B, \infty) \to \mathbb{R}$ by splicing at B the low-action ODE solution L^* with the high-action ODE solution F^*. That is, let

$$v(W) \equiv \begin{cases} L^*(W) & \text{for} \quad W = B, \\ F^*(W) & \text{for} \quad W > B. \end{cases} \tag{17}$$

Theorem 1 (Verification) *Suppose* $B \in (0, \bar{\boldsymbol{B}}]$ *and* $W_0 \in [B, W_{gp}(B)]$. *Then* $v(W_0) = V(W_0)$, *i.e.,* v *is the firm's value function in the contracting problem with the agent's outside option* B *and the initial value* W_0. *The optimal controls* c, a, Y *attaining* v *define an optimal contract with* $C_t = c(W_t)$, $A_t = a(W_t)$, *and* $Y_t = Y(W_t)$, *where* $\{W_t; 0 \le t < \infty\}$ *is a weak solution to* (7), *with stopping times* $\tau_{tn} = \infty$ *and* $\tau_{gp} = \min\{t : W_t = W_{gp}(B)\}$.

Proof Follows by setting $R = 0$ in Theorem 2. ■

3.5 Reflective dynamics of the optimal contract

Similar to Zhu (2013), the suspension phase of the optimal contract generates an upward reflection of the agent's continuation value process W_t at the lower bound B. Indeed, in the optimal contract of Theorem 1, policies c, a, and Y are taken from the low-action solution L^* when $W_t = B$, i.e., $c(B) = a(B) = Y(B) = 0$. Thus, by (15), W_t moves upward deterministically (with no volatility) whenever $W_t = B$. With sample paths of W_t being continuous, W_t can never drop below B, i.e., (2) holds. Also, using $W_t = B$ on the right side of (15), we have $dW_t = rBdt$, i.e., the drift of W_t at the lower bound B is stronger (more positive) when B is higher.

Furthermore, since the high-action solution F^* has $Y(W) \ge \gamma_0 > 0$ for all $W \in (B, W_{gp}(B))$, the process W_t is similar to sticky Brownian motion around its lower bound B, as in Zhu (2013). After hitting B, W_t moves out of B immediately but then returns to B frequently. As a consequence of these frequent revisits, although each visit to B has zero duration, the total expected amount of time that W_t spends at B is strictly positive. This kind of reflection is known as slow reflection in the literature (see, e.g., Harrison and Lemoine (1981), Bou-Rabee and Holmes-Cerfon (2020)).

4 Termination versus suspension with positive residual value

In this section, we study how the project's residual value, R, affects the use of termination and suspension in an optimal contract. The project's residual value could be coming from replacing the agent and continuing the project's operation, as in Sannikov (2008), or from liquidating the project, as in DeMarzo and Sannikov (2006) or Zhu (2013).

The residual value R is the opportunity cost of suspension, as the firm passes on the option to collect R whenever it chooses to suspend the agent instead of terminating. With higher R, naturally, the firm is more willing to terminate or retire the agent. The agent's outside option B, however, has a nonmonotonic impact on the trade-off between termination and suspension.

In Section 4.2, we derive a condition that determines if, for a given B and R, suspending the agent at B dominates the option to terminate the relationship and collect R. In Section 4.4, in Proposition 2, we characterize the set of pairs (B, R) for which this condition is met. Section 4.6 illustrates the impact of B and R on the optimal contract using three numerical examples. Sections 4.1, 4.3, and 4.5 are more technical, as they deal with, respectively, the boundary conditions, the existence and classification of high-action ODE solutions, and verification of the optimality of suspension.

Due to the firm's free disposal of its project, we restrict attention to $R \geq 0$. On the high end, we restrict attention to R that satisfy a loose upper bound given in equation (24) in Appendix A.1 because the relationship will not form, i.e., the agent will not be hired, if R exceeds this bound.

4.1 Boundary conditions for termination and suspension

To construct optimal contacts for various levels of the firm's residual value R, we will use the same approach as in Section 3, which combines solutions to the low- and high-action ODEs, (10) and (11). The two ODEs are independent of R. In particular, the low-action ODE solutions, given in (14), are unaffected by R. However, since the residual value R enhances the firm's profit at agent termination and retirement, the level of R matters for the boundary conditions used with the high-action ODE at termination, retirement, and suspension.

Generalizing (12), the firm's payoff upon terminating the agent at $W = B$ is

$$F(B) = R. \tag{18}$$

Generalizing (13), the agent's retirement threshold, W_{gp}, is determined by the requirement $F(W) \geq R + F_0(W)$ for all $W \geq B$, and the smooth-pasting conditions

$$F(W_{gp}) = R + F_0(W_{gp}) \quad \text{and} \quad F'(W_{gp}) = F_0'(W_{gp}). \tag{19}$$

An optimal contract with termination at B, as in Section 3.1, is obtained from a unique high-action solution satisfying (18) and (19). As before, we will denote this solution by $\tilde{F}$ and refer to it as the firm's profit function with termination at B. As in Lemma 1, two cases are possible. If the initial slope of this profit function, $\tilde{F}'(B)$, is strictly positive, the optimal terminating contract starts at $W_0 = \text{argmax}_W \tilde{F}(W) > B$ and ends when the agent's continuation value exits the interval (B, W_{gp}). If the initial slope of $\tilde{F}$ is not strictly positive, the firm does not offer a contract to the agent but rather collects R immediately.

Similarly, generalizing (16), suspension of the agent at $B > 0$ requires smooth splicing between high- and low-action ODE solutions at $W = B$ and at some level $R + y$. The low-action ODE solution L through the point $(B, R + y)$ has slope $(R + y)/B$. The boundary conditions for smooth splicing at that point, therefore, are

$$F(B) = L(B) = R + y \quad \text{and} \quad F'(B) = L'(B) = \frac{R + y}{B}. \tag{20}$$

4.2 Criterion for the optimal use of suspension

A critical test for whether suspension after poor agent performance dominates termination comes from comparing the slopes of the low-action solution L and the termination profit function $\tilde{F}$ at (B, R). If L is flatter than $\tilde{F}$ at (B, R), we can find a positive boundary premium y with which a high-action solution F exists that satisfies both the optimal retirement condition (19) and the condition for smooth splicing with L at the point $(B, R + y)$, (20). This F lies above the termination profit function $\tilde{F}$ and, hence, suspension of the agent at B dominates termination. If L is steeper than $\tilde{F}$ at (B, R), however, no such F exists and, hence, termination at B is optimal.

That the ranking of the slopes of L and $\tilde{F}$ is critical for the use of suspension versus termination at B can be seen directly from the following approximation. Suppose the agent's suspension has to last $\triangle$ units of time, where $\triangle$ is small but strictly positive. We want to see if the firm can benefit from suspending instead of terminating when $W_t = B$, i.e., if such a deviation from $\tilde{F}$ can yield higher profit at B than $\tilde{F}(B) = R$. From the law of motion (15), we know that the agent's continuation value at the end of suspension will be $W_{t+\triangle} = e^{r\triangle} W_t = e^{r\triangle} B$. The firm's profit at the end of suspension, thus, will be $\tilde{F}(e^{r\triangle} B)$. Since the firm's expected profit flow is zero during suspension (no effort, no compensation), the profit at the start of suspension is simply $e^{-r\triangle} \tilde{F}(e^{r\triangle} B)$, which to a first-order approximation is $\tilde{F}(B) + (\tilde{F}'(B)B - \tilde{F}(B))r\triangle$. This profit dominates the value of termination at B, $\tilde{F}(B)$, if and only if $\tilde{F}'(B)B - \tilde{F}(B) \geq 0$, i.e.,

$$\tilde{F}'(B) \geq \frac{\tilde{F}(B)}{B} = \frac{R}{B}. \tag{21}$$

Since R/B is the slope of the low-action solution L passing through the point (B, R), the above condition is equivalent to

$$\tilde{F}'(B) \geq L'(B),$$

i.e., suspension dominates termination if and only if L is flatter at (B, R) than $\tilde{F}$.

Condition (21) captures the cost and benefit of suspending the agent at B. The right side of (21) represents the total cost of suspension: the flow opportunity cost R times the duration of suspension $1/B$. The firm's outside option, R, is equivalent to receiving the flow profit R forever: $R = \int_0^\infty re^{-rt}Rdt$. During suspension, the firm's profit flow is zero, i.e., less than R. Recall from Section 3.5 that the reflection of W_t off B is slow and, hence, the total duration of suspension is positive. Recall from equation (15) that the drift of W_t at B is rB. The factor $1/B$ on the right side of (21), thus, represents the duration of suspension. The left side of (21) represents the benefit of using suspension. Since $\tilde{F}(B) = R$, the slope $\tilde{F}'(B)$ shows how much more profitable than termination the relationship becomes if suspension is used at B instead of terminating.

If $R = 0$, then, as shown in Section 3, for any $B \in (0, \bar{\boldsymbol{B}}]$, the low-action solution L is flatter at the point $(B, R) = (B, 0)$ than the termination profit function $\tilde{F}$, i.e., condition (21) is met, simply because the slope of L is $R/B = 0$ and $\tilde{F}'(B) \geq 0$ for any $B \leq \bar{\boldsymbol{B}}$. Clearly, since suspension carries no opportunity cost when $R = 0$, suspension dominates termination whenever suspension is feasible (i.e., $B > 0$) and the relationship is profitable (i.e., $B \leq \bar{\boldsymbol{B}}$). If $R > 0$ and $B > 0$, however, whether condition (21) is met depends on the values of R and B. We study this question in Section 4.4.

4.3 Classification of solution curves

In this section, we show the existence and uniquness of a solution to the high-action ODE equation with a fixed boundary condition at B and a free-boundary condition pinning down the point of the agent's retirement. We show the continuity and monotonicy of the initial slope of this solution with respect to its boundary value.

In Appendix A.1, we define a set $\mathcal{N}$ such that it is optimal for the firm to not run its project if the outside options (B, R) and the boundary premium y are outside of $\mathcal{N}$. We restrict attention to (B, R, y) that belong to the closure of $\mathcal{N}$, $cl(\mathcal{N})$.

Lemma 2 *i. For each $(B, R, y) \in cl(\mathcal{N})$, there exists a unique solution F to the high-action ODE, (10), satisfying $F(B) = R + y$, $F(W) \geq R + F_0(W)$ for all $W \geq B$, and the smooth-pasting conditions (19) at some $W_{gp} \in [\max\{B, 0\}, W^*_{gp}]$. The solution F is strictly concave if $(B, R, y) \in \mathcal{N}$.*

ii. Denote the initial slope of F, $F'(B)$, as $x(B, R, y)$. The function $x(B, R, y)$ is continuous on $\mathcal{N}$.

iii. If $B > 0$, then $x(B, R, y)$ is strictly decreasing in both R and y.

Proof In Appendix A.3. ■

Sannikov (2008) uses a forward-shooting procedure to pin down a unique solution to the high-action ODE that satisfies a level condition at the left boundary, B, and pastes smoothly with F_0 at an endogenous right boundary, W_{gp}. In this procedure, if a candidate solution remains everywhere strictly above F_0, its initial slope is too high; if a candidate solution crosses F_0, its initial slope is too low. The first part of Lemma 2 verifies that for each $(B, R, y) \in \mathcal{N}$ the same procedure pins down a unique solution F to the high-action ODE (10), starting from the initial level $F(B) = R + y$ and pasting smoothly with $R + F_0$.

In the third part of Lemma 2, it is obvious that a higher boundary premium y forces the initial slope of the solution F, $x(B, R, y)$, to be lower, for otherwise F would remain strictly above $R + F_0$. It is not obvious, however, that $x(B, R, y)$ should also be decreasing in R, as both the initial level $F(B) = R + y$ and the retirement payoff curve $R + F_0$ increase uniformly with R. The intuition for why this is the case comes from the fact that higher R reduces the firm's aversion to the risk of early termination. This aversion is captured by the second derivative of F, and higher R makes solutions F less concave.[25] Starting from $F(B) = R+y$, a less-concave solution curve is more likely to stay above $R + F_0$, which means the solution curve that goes down to $R + F_0$ (and pastes with it smoothly) must have a lower initial slope.

4.4 Regions of termination and suspension

For each $(B, R, y) \in \mathcal{N}$, we want to know if $x(B, R, y)$ is positive because otherwise, as we saw in Section 3, the optimal course of action is to not offer a contract but rather to collect R without delay. Furthermore, we want to know if $x(B, R, 0)$ is larger than R/B, the slope of the low-action ODE solution through the point (B, R), because in these cases, as shown in (21), suspension at B gives rise to a positive boundary premium y, thus dominating termination.

Proposition 2 *i. **(Region of no contract)** For each $R \in [0, \bar{A})$, there exists a $\bar{B}(R) \geq 0$ such that*

$$x(B, R, 0) \begin{cases} > 0, & \text{if } B < \bar{B}(R); \\ = 0, & \text{if } B = \bar{B}(R); \\ < 0, & \text{if } B > \bar{B}(R). \end{cases} \tag{22}$$

[25] Mechanically, as higher R raises both boundary values $F(B)$ and $F(W_{gp})$, it also raises $F(W)$ for all $W \in (B, W_{gp})$. In (25) in the Appendix, $H_a(W, F(W), F'(W))$ is decreasing in $F(W)$, i.e., higher $F(W)$ makes $F''(W)$ less negative.

In particular, $\bar{B}(0) = \bar{\boldsymbol{B}}$*, where* $\bar{\boldsymbol{B}}$ *is defined in Lemma 1.*

There exists $\bar{\boldsymbol{R}} \in (0, \bar{A})$ *such that* $\bar{B}(R) \begin{cases} > 0, & \textit{if } R \in [0, \bar{\boldsymbol{R}}); \\ = 0, & \textit{if } R \in [\bar{\boldsymbol{R}}, \bar{A}). \end{cases}$

On $[0, \bar{\boldsymbol{R}}]$*, the function* $\bar{B}(R)$ *is continuous and strictly decreasing. Let* $\bar{R}(B) : [0, \bar{\boldsymbol{B}}] \to [0, \bar{\boldsymbol{R}}]$ *denote the inverse of* $\bar{B}$ *on* $[0, \bar{\boldsymbol{B}}]$*. In particular,* $\bar{R}(0) = \bar{\boldsymbol{R}}$ *and* $\bar{R}(\bar{\boldsymbol{B}}) = 0$*. The function* $\bar{R}(B)$ *is continuous and strictly decreasing.*

ii. ***(Regions of termination and suspension)*** *For any* $B \in (0, \bar{\boldsymbol{B}}]$*, there exists a unique* $R^*(B) \in [0, \bar{R}(B)]$ *such that*

$$x(B, R, 0) - \frac{R}{B} \begin{cases} > 0, & \textit{if } R \in [0, R^*(B)); \\ = 0, & \textit{if } R = R^*(B); \\ < 0, & \textit{if } R \in (R^*(B), \bar{R}(B)). \end{cases}$$

If $B \in (0, \bar{\boldsymbol{B}})$*, then* $R^*(B) > 0$*. If* $B = \bar{\boldsymbol{B}}$*, then* $R^*(B) = 0$*. In particular,* $R^{*\prime}(0) > 0$ *and* $R^{*\prime}(\bar{\boldsymbol{B}}) < 0$*.*

iii. ***(Positive boundary premium from suspension)*** *If* $B \in (0, \bar{\boldsymbol{B}}]$ *and* $R \leq R^*(B)$*, i.e.,* $R/B \leq x(B, R, 0)$*, then there exists a unique* $y^* \geq 0$ *such that* $x(B, R, y^*) = (R + y^*)/B$*. If* $R = R^*(B)$*, then* $y^* = 0$*. If* $R < R^*(B)$*, then* $y^* > 0$*.*

Proof In Appendix A.4. ■

Proposition 2 identifies two critical boundaries, $\bar{R}(B)$ and $R^*(B)$, that separate all pairs $(B, R) \geq (0, 0)$ into three regions. For (B, R) high enough, i.e., $B \geq \bar{\boldsymbol{B}}$ or $B < \bar{\boldsymbol{B}}$ and $R \geq \bar{R}(B)$, we have $x(B, R, 0) = \tilde{F}'(B) < 0$, i.e., the termination profit function $\tilde{F}$ initiated at (B, R) is monotonically decreasing, which means the option of collecting R immediately dominates the option of offering a contract to the agent.[26]

In the middle region, i.e., $B < \bar{\boldsymbol{B}}$ and $R^*(B) < R < \bar{R}(B)$, the termination profit function $\tilde{F}$ is initially upward-sloping, which means the optimal contract starting at $W_0 = \operatorname{argmax}_W \tilde{F}(W) > B$ and terminating at B dominates the firm's outside option R. Suspension at B cannot generate a positive boundary premium y because condition (21) is violated. Specifically, $R/B > x(B, R, 0) = \tilde{F}'(B)$ implies that no high-action ODE solution exists that satisfies the boundary conditions (19) and (20) with a positive boundary premium $y \geq 0$.[27] For (B, R) in this region, termination at B is optimal.

[26] Reflecting the impact of moral hazard, the boundary $\bar{R}(B)$ lies strictly below the upper bound in inequality (24), which is derived in Appendix A.1 in the absence of moral hazard.

[27] Indeed, with $y = 0$, the high-action ODE solution F with boundary conditions $F(B) = L(B) = R$ and $F'(B) = L'(B) > \tilde{F}'(B)$ stays above F_0 for all $W > B$, so the retirement condition (19) is not met. A strictly positive $y > 0$ will shift F further upward by increasing both its initial level, $R + y$, and slope, $(R + y)/B$. A contract with suspension at B can be constructed with some negative boundary premium $y < 0$ such that $R > R + y > 0$, which only confirms that termination dominates suspension at B.

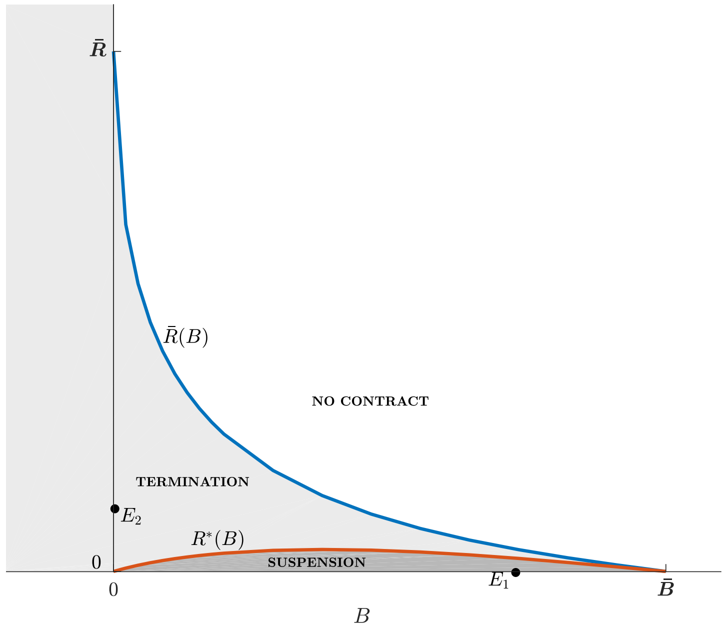

Figure 2: Regions of no contract, termination, and suspension in the plane (B, R). Parameter values as in Figure 1. Example $E_1 = (B_1, R_1)$ has $B_1 = 0.9W_0$, where W_0 is the agent's value at the start of a contract, and $R_2 = 0$. Example $E_2 = (B_2, R_2)$ has $B_2 = 0$ and $R_2 = 0.9v_2(W_0)$, where $v_2(W_0)$ is the firm's value at the start of a contract.

In the bottom region, $B < \bar{\boldsymbol{B}}$ and $R \leq R^*(B)$, condition (21) is met, which means suspension can generate a positive boundary premium $y \geq 0$. Indeed, the low-action ODE solution is flatter at B than the terminating high-action ODE solution $\tilde{F}$, i.e., with $y = 0$ we have $R/B \leq x(B, R, 0) = \tilde{F}'(B)$. As we increase y above 0, similar to Proposition 1, the slope $(R + y)/B$ increases and the slope $x(B, R, y)$ decreases. We then find $y^* \geq 0$ such that $(R + y^*)/B = x(B, R, y^*)$. With this level of the boundary premium, we have a unique high-action ODE solution, F^*, such that the smooth splicing condition (20) holds:

$$F^*(B) = R + y^* \text{ and } F^{*\prime}(B) = x(B, R, y^*) = \frac{R + y^*}{B}. \tag{23}$$

In fact, $y^* > 0$ for all $R < R^*(B)$. This bottom region includes the case of $R = 0$ already discussed in Section 3. We formally verify in the next section that an optimal contract, which suspends the agent at B, can be constructed from the solution F^*.

Suspension cannot be implemented at any $B \leq 0$ because the agent's flow value of suspension, $u(0) - h(0) = 0$, is above B in these cases, i.e., suspension could only push the agent's continuation value down, not up, which would violate (2). Termination at B, thus, is optimal for all $B \leq 0$. Specifically, if $B = 0$, the termination profit function $\tilde{F}$ has a strictly positive initial slope if $R < \bar{\boldsymbol{R}}$ and zero initial slope if $R \geq \bar{\boldsymbol{R}}$, i.e., it is optimal to not offer a contract if the project's residual value is sufficiently high.[28] If $B < 0$, the termination profit function $\tilde{F}$ has a strictly positive initial slope for all $R \geq 0$, i.e., a nondegenerate terminating contract is optimal.[29]

Figure 2 illustrates the results of Proposition 2 using a numerical example with the same parameter values as in Figure 1. The relationship is not profitable, i.e., no contract is offered to the agent, if (B, R) belongs to the unshaded region. The boundary of profitability of the relationship, $\bar{R}(B)$, is convex, which shows that the two parties' outside options, R and B, reinforce each other in reducing the value of the relationship. Termination upon the first visit of W_t to B is optimal if (B, R) belongs to the light-shaded region. In the dark-shaded region, suspension at B is optimal. This region contains many economically relevant cases, in which the firm's outside option is low and the agent's outside option is moderate.[30] The boundary of the region of optimality of suspension, $R^*(B)$, is hump-shaped in this example.

The shape of the boundary R^* is determined by the two opposing effects that the agent's outside option B has on the value of using suspension versus termination. On the one hand, higher B increases the drift of W_t during suspension, as in (15), which reduces the total duration of suspension, making suspension less costly, i.e., more useful. On the other hand, higher B makes the relationship outside of suspension less profitable, as it tightens the agent's participation constraint, (2), which makes suspension less useful. In condition (21), the first effect is captured by $1/B$ and the second by $\tilde{F}'(B)$.

At low B, the first effect dominates, which means $R^*(B)$ is increasing. Indeed, an increase in B in this area reduces the duration of suspension by a lot while reducing the profitability of the relationship out of suspension only by a little, which increases the overall value of suspension. Thus, the level of R under which the firm remains indifferent between suspending and terminating, $R^*(B)$, must increase. Note, in particular, that if $B = 0$, then the drift of W_t in suspension is zero, i.e., the duration of suspension is infinite, which makes suspension not useful at all. Thus, $R^*(0) = 0$. At high B, by contrast, the second effect dominates. There, an increase in B does not reduce the duration of suspension strongly, but it continues to reduce the profitability of the relationship out of suspension, pushing it down to zero at

[28] With $R \geq \bar{\boldsymbol{R}}$, the curvature of F is so low that, even with $F'(0) = 0$, F stays strictly above $R + F_0$ for all $W > 0$, i.e., $W_{gp} = 0$. Sannikov (2008) also finds cases with low curvature of F coming from a high discount rate r or a high volatility σ. Throughout our analysis, we exclude these cases, i.e., we assume that r and σ are not so high to imply $W_{gp} = 0$ under $B = R = 0$. In particular, this means that $x(0, 0, 0) > 0$.

[29] Indeed, with $B < 0$, the agent strictly prefers to be retired with the retirement value $W_{gp} = 0$ to being fired, while the firm is indifferent to these two outcomes at contract completion. At least for a short while, thus, the agent can be incentivized to exert maximum effort $\bar{A}$ just via the promise of retirement with $W_{gp} = 0$ after good performance, without any other compensation. Since $\bar{A} > R$ for all R in $\mathcal{N}$, this effort sufficiently compensates the firm for delaying its collection of R.

[30] In Remark 1, these cases correspond to the firm facing relatively high costs in the process of searching for the agent's replacement and the agent facing moderate costs in the process of searching for a new job.

$B = \bar{B}(R) \leq \bar{\boldsymbol{B}}$.

Consistent with the intuition provided by this example, the second part of Proposition 2 shows that the boundary $R^*(\cdot)$ is always initially increasing and eventually decreasing. We do not have a proof that $R^*(\cdot)$ is always single-peaked, but neither do we have a counterexample.

4.5 Optimal contract: formal statement and verification

For (B, R) such that $B \in (0, \bar{\boldsymbol{B}}]$ and $R \leq R^*(B)$, let $y^*(B, R)$ denote the boundary premium y^* identified in the third part of Proposition 2, let F^* denote the high-action ODE solution that satisfies (23), and let $W_{gp}(B, R)$ denote the associated retirement threshold W_{gp}. Let $L^*(W) = (R + y^*(B, R))B^{-1}W$ denote the unique low-action ODE solution through the point $(B, R + y^*(B, R))$. With these F^* and L^*, define $v : [B, \infty) \to \mathbb{R}$ as in (17).

Theorem 2 (Verification) *Suppose $B \in (0, \bar{\boldsymbol{B}}]$, $R \leq R^*(B)$, and $W_0 \in [B, W_{gp}(B, R)]$. Then $v(W_0) = V(W_0)$, i.e., $v(W_0)$ is the firm's value function in the contracting problem with the agent's outside option B, the firm's residual value R, and the agent's initial value W_0. The optimal controls c, a, Y attaining v define an optimal contract with $C_t = c(W_t)$, $A_t = a(W_t)$, and $Y_t = Y(W_t)$, where $\{W_t; 0 \leq t < \infty\}$ is a weak solution to (7), with stopping times $\tau_{tn} = \infty$ and $\tau_{gp} = \min\{t : W_t = W_{gp}(B, R)\}$.*

Proof In Appendix A.6 ■

The proof follows Sannikov (2008) very closely with two exceptions. The technical argument for the existence of a solution to (7) is modified to account for volatility of W_t vanishing at B, and the step verifying the optimality of the contract is modified to account for the reflection of the process W_t at B and $\tau_{tn} = \infty$. For pairs (B, R) for which termination at B is optimal, the statement and verification of the optimal contract follows Theorem 3 in Sannikov (2008) with no significant changes.

4.6 The impact of B and R on the optimal contract

In this section, we illustrate the impact of the agent's outside option B and the firm's residual value R on the optimal contract using three numerical examples. As our baseline, we take the main example of Sannikov (2008), where $B = R = 0$. We compare the optimal contract from this baseline against optimal contracts obtained in two examples. Example E_1 has a relatively high B, and example E_2 has a relatively high R. In example E_1, the agent's outside option satisfies $B_1 = 0.9W_0 \in (0, \bar{\boldsymbol{B}})$, where $W_0 = \text{argmax}_{W \geq B_1} v_1(W)$, and the firm's residual value is $R = 0$. In example E_2, $B = 0$ and $R_2 = 0.9v_2(W_0) \in (0, \bar{\boldsymbol{R}})$, where $W_0 = \text{argmax}_{W \geq 0} v_2(W)$.

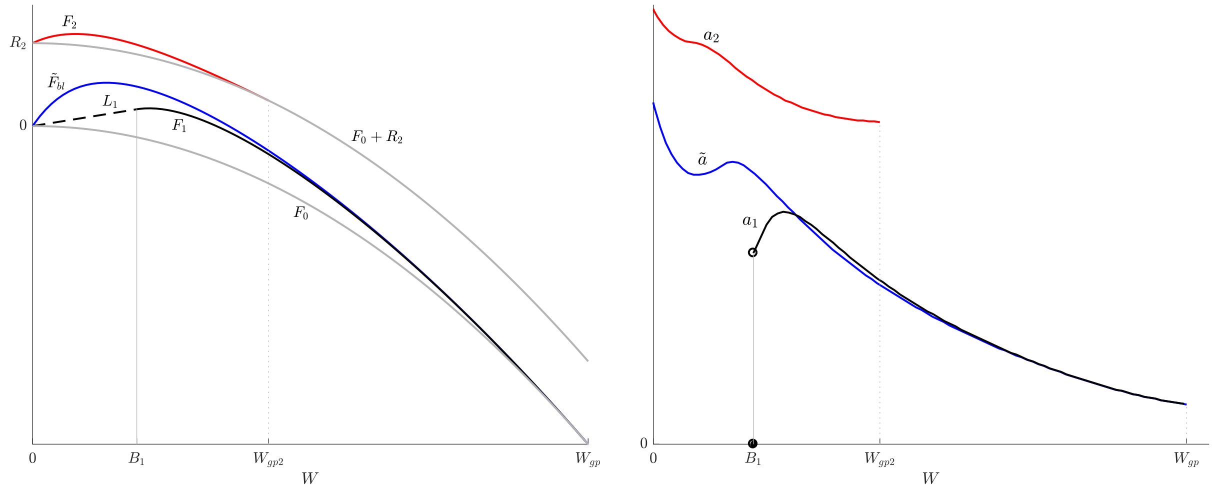

Figure 3: Optimal solution curves and their associated effort policy functions in three examples. Baseline example with $B = R = 0$: solution curve $\tilde{F}_{bl}$ (left panel) and its effort policy function $\tilde{a}$ (right panel). Example E_1 with $B = B_1 > 0$ and $R = 0$: solution L_1 spliced with F_1, and the associated effort policy function a_1 (discontinuous at B_1). Example E_2 with $B = 0$ and $R = R_2 > 0$: solution F_2 and its effort policy a_2. Parameter values as in figures 1 and 2.

In equation (1), thus, example E_1 corresponds to $\kappa_a = 0.1$ and $\kappa_f = 1$, while E_2 corresponds to $\kappa_a = 1$ and $\kappa_f = 0.1$. In Figure 2, the pairs (B, R) for these two examples are marked as points E_1 and E_2, respectively. Other than B and R, parameters used to compute the optimal contract in the baseline and in the two examples are the same as in figures 1 and 2.

Figure 3 shows the optimal low- and high-action ODE solution curves used to construct the respective value functions v in the three examples (left panel) along with their associated optimal effort policy functions a (right panel). Relative to baseline, the firm achieves a lower profit in example E_1, where the agent's participation constraint (2) is tighter. Consistent with Section 3, the optimal contract in E_1 is qualitatively different from the baseline: termination of the agent after poor performance is not optimal. It is worth pointing out that, despite the retirement threshold W_{gp} in E_1 being very close to that of the baseline, the ex ante expected duration of the optimal contract is much longer in E_1 because the contract does not terminate at the lower bound but only exits at the retirement threshold W_{gp}.

In example E_2, the optimal contract is qualitatively the same as in baseline: it terminates the agent after sufficiently poor performance and retires him after strong performance. Although the firm's ex ante profit is much higher than in the baseline, most of this value comes from the residual value R_2 itself. The agent's value, W_0, is lower than in baseline, while his effort, marked as a_2 in the right panel of Figure 3, is higher. The retirement threshold, W_{gp2}, is much lower than the retirement threshold W_{gp} in the baseline. Consistent with high volatility of the agent's continuation value implied by high effort a_2 as well as with the low retirement threshold W_{gp2}, the ex ante expected duration of the optimal contract in example E_2 is very short relative to the baseline.

It is worth pointing out that the low expected duration of the contract in example E_2 follows from the fact that the firm collects R_2 only after the completion of the contract. With R_2 being relatively high, the firm desires a quick termination of the contract in order to avoid a long delay in its collection of this value. With quick termination desirable, the firm is less averse to volatility in the agent's continuation value W_t, as higher volatility increases the chance of reaching either of the contract exit points, 0 or W_{gp2}, quickly. Consequently, the agent is exposed to steep incentives and supplies high effort. Intuitively, we can say that high residual value R makes the firm more impatient and less risk averse, which makes the contract's duration short and the agent's effort high.

5 Renegotiation

Thus far, we have assumed the contracting parties' ability to commit at $t = 0$ to not renegotiate the terms of the contract at any future date. This assumption is binding whenever the resulting profit function $v(W)$ is hump-shaped, as both parties would benefit from a one-time shift of the contract to the peak of v as soon as the agent's continuation value W_t enters the region in which v is upward-sloping. If v is nonincreasing, the resulting contract is renegotiation-proof (RP).

In this section, we briefly discuss how the requirement of renegotiation-proofness changes the optimal contract. In addition to the standard RP contract of DeMarzo and Sannikov (2006) with stochastic termination, we show that a suspension contract is RP if $R = 0$.

In our model, the standard RP contract of DeMarzo and Sannikov (2006) is optimal in all cases in which the relationship is viable, i.e., for all $R \geq 0$ and $B \leq \bar{B}(R)$. This contract uses stochastic termination at $\bar{B}(R)$. In particular, let c, a, Y denote the policy functions that in the high-action ODE (10) attain the termination profit function $\tilde{F}$ initiated at $(\bar{B}(R), R)$. Note that, by the first part of Proposition 2, this $\tilde{F}$ is downward-sloping with $\tilde{F}'(\bar{B}(R)) = 0$. In the RP contract of DeMarzo and Sannikov (2006), the agent's continuation value W_t evolves on the interval $[\bar{B}(R), W_{gp})$ according to

$$dW_t = r(W_t - u(c(W_t)) + h(a(W_t)))dt + rY(W_t)(dX_t - a(W_t)dt) + dP_t,$$

where P_t is an increasing process that satisfies $(W_t - \bar{B}(R))dP_t = 0$ at all t.[31] This law of

[31] By the first part of Proposition 2, the interval $[\bar{B}(R), W_{gp})$ is nonempty for all $R < \bar{\boldsymbol{R}}$. For $R \geq \bar{\boldsymbol{R}}$, we have

motion for W_t is the same as (4) whenever $W_t > \bar{B}(R)$. At $\bar{B}(R)$, the agent is either terminated, in which case his continuation value jumps down to $B \leq \bar{B}(R)$, or his continuation value is reflected upward (i.e., W_t is increased by $dP_t > 0$) and the relationship continues. The ex ante probability that the agent is not terminated by time t is $\exp\left(-P_t/(\bar{B}(R) - B)\right)$. The reflection of W_t at $\bar{B}(R)$ achieved here by the process P_t is stochastic, i.e., dP_t is correlated with dX_t, and fast, i.e., the optimal RP contract spends zero total time at $\bar{B}(R)$.

In addition to this standard contract with stochastic termination, if $R = 0$, the optimal contact with suspension of the agent at $\bar{B}(0) = \bar{\boldsymbol{B}}$ is another optimal RP contract. On the interval $(\bar{\boldsymbol{B}}, W_{gp})$, this contract uses the same policies as the standard contract with stochastic termination. When $W_t = \bar{\boldsymbol{B}}$, however, it uses suspension instead of stochastic termination. Since this contract never terminates the agent (i.e., $\tau_{tn} = \infty$), the outside value B is never used to deliver the continuation value W_t to the agent, which makes this contract feasible for any $B \leq \bar{\boldsymbol{B}}$. It follows that the optimal contact with suspension of the agent at $\bar{\boldsymbol{B}}$ is an optimal RP contract for $R = 0$ and any $B \leq \bar{\boldsymbol{B}}$.[32]

6 Conclusion

In this paper, we examine boundary behavior of optimal contracts in a standard dynamic principal-agent model with moral hazard. We find that existing literature overemphasizes the necessity of terminating the relationship when the agent's stake in the relationship runs out, i.e., when the agent's binding participation constraint implies that standard pay-for-performance incentives must be switched off. Rather, we find temporary suspension of the agent to be a feasible and efficient alternative to termination in a robust set of cases.

We examine how the trade-off between termination and suspension depends on the firm and the agent's outside options. A higher firm outside option, predictably, increases the firm's desire to terminate. Specifically, we show that a higher outside option makes the firm more impatient and less risk averse, which increases both agent effort and turnover. A higher agent outside option has a nonmonotonic impact on this trade-off, as it simultaneously increases the efficiency of suspension by making suspension shorter and decreases the firm's desire to suspend by reducing the overall profitability of the relationship. We show that this trade-off is captured by a condition relating the slope of the firm's profit function under the assumption of termination to the ratio of the outside options of the firm and the agent. This condition provides a simple test to determine the optimal boundary behavior of a contract.

$\bar{B}(R) = W_{gp} = 0$, and the optimal RP contract retires the agent immediately at $t = 0$.

[32] Indeed, with $R = 0$, all high-action ODE solutions F with $F(B) > 0$ and $F'(B) \leq 0$ violate both the termination boundary condition (18) and the suspension boundary condition (20) for all $y \geq 0$.

Our analysis can be extended to examine additional interesting optimal contracting questions. First, our standard model has only one action free of moral hazard, the action of exerting no effort used in the suspension phase of the contract. If the set of moral-hazard-free tasks is richer, we can ask which of these tasks produces the most efficient suspension. The trade-off would involve the productivity of the task and the disutility it causes to the agent, with both being desirable in agent suspension. Second, we can examine how other contract-exit possibilities, in addition to the firing or retiring of the agent, would affect the trade-off between termination and suspension. If the firm has more flexibility in separating from the agent after strong performance, for example by combining the agent's outside option with a severance payment, the value of suspending the agent after poor performance increases while the value of terminating at that point remains the same. Such added flexibility would therefore increase the incidence of suspension. Third, the intrinsic motivation of the agent, where the agent always provides at least some minimum level of effort, can be captured as an additional constant term in the high-action ODE equation to show that, ceteris paribus, intrinsic motivation enhances the use of suspension. Furthermore, by imposing more structure in a model embedding our contracting problem into an external labor market, our analysis can be extended to study the links between labor market search or bargaining frictions and the optimal use of termination as an incentive device under moral hazard.

Appendix

A.1 An upper bound on (B, R)

If outside option values B and R are sufficiently high, it is optimal for the firm to not start its project but rather to collect its residual value R immediately. Based on this observation, in this section, we derive an upper bound on B and R above which not starting the project, i.e., not offering a contract to the agent, is optimal. This bound is loose because we assume that the agent's action is observable and contractible, i.e., moral hazard is absent, in this section.

Suppose the firm does not run the project but rather collects its residual value R immediately. Without running the project, the firm can still deliver any value $W \geq B$ to the agent by retiring him, terminating him, or using a lottery between retirement and termination. Define $F_{\text{cav}}(W)$ as the concavification on the half-line $W \geq B$ of the firm's payoff from retiring the agent with value W, which pays $R + F_0(W)$ to the firm, and terminating him with value B, which pays R to the firm. For any B and $R \geq 0$, there exists a unique straight line through the point (B, R) that is tangent to $R + F_0(\cdot)$. Denote by $T(B)$ the horizontal coordinate of the

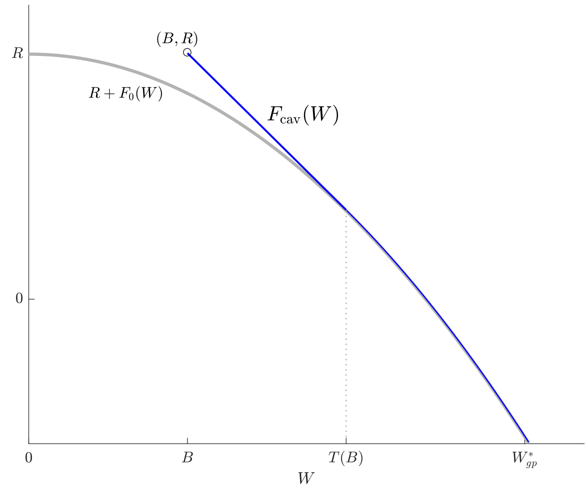

Figure 4: The construction of F_{cav}. The straight segment connecting (B, R) and $(T(B), R+F_0(T(B)))$ has slope $F_0'(T(B))$.

point of tangency.[33] The concavification of the firm's payoff from retirement or termination of the agent is given by

$$F_{\text{cav}}(W) \equiv \begin{cases} R + F_0(T(B)) + F_0'(T(B))(W - T(B)), & \text{if } W \in [B, T(B)]; \\ R + F_0(W), & \text{if } W \geq T(B). \end{cases}$$

Since $F_{\text{cav}}(W)$ is decreasing and concave, its slope is the least negative at $W = B$, where it is equal to the slope of the tangent line through (B, R), i.e., $F_{\text{cav}}'(B) = F_0'(T(B))$. Figure 4 provides an illustration.

Now suppose the firm considers delaying its collection of R by an instant and asking for some positive effort a from the agent during this short spell. In order to compensate the agent for the effort a, the firm needs to increase his W by $h(a)$.[34] Under the value function F_{cav}, the cost to the firm of increasing the agent's value by $h(a)$ is $-F_{\text{cav}}'(W)h(a)$, where $-F_{\text{cav}}'(W)$ represents the marginal cost of delivering utility to the agent. This cost is weakly larger than $-F_{\text{cav}}'(B) = -F_0'(T(B))$.

The profit gain (in flow terms) resulting from such a deviation from $F_{\text{cav}}(W)$ is at most

$$-R + \max_{a \in [0, \bar{A}]} \{a + F_0'(T(B))h(a)\},$$

[33] Note that $T(B) \geq B$ is independent of R, equal to zero for all $B \leq 0$, and strictly increasing in B for all $B > 0$. For example, if $u(c) = \sqrt{c}$, then $T(B) = \max\{2B, 0\}$.

[34] This is simply compensation for the disutility of effort. There are no additional incentive costs because we assume the absence of moral hazard in this section.

where the first term is the flow cost of delaying the collection of R, and the term under maximization is the net gain from having the agent exert positive effort, assuming the cost of compensating the agent for effort is the lowest possible, $-F_0'(T(B))$. If this profit gain is nonpositive, then it is optimal for the firm to not run its project but rather collect its residual value R immediately, even in absence of moral hazard. We can therefore restrict attention to pairs (B, R) that satisfy

$$R < \max_{a\in[0,\bar{A}]} \{a + F_0'(T(B))h(a)\}. \tag{24}$$

In particular, with $R = 0$ the above inequality is equivalent to $T(B) < W_{gp}^*$, where W_{gp}^* solves $F_0'(W) = -1/h'(0)$. Indeed, using the first-order condition with respect to a, it is easy to check that the right side of (24) is positive if and only if $1 + F_0'(T(B))h'(0) > 0$. Thus, with $R = 0$ inequality (24) is equivalent to

$$B < \hat{B},$$

where $\hat{B} \equiv T^{-1}(W_{gp}^*)$.[35] Note also that (24) rules out $R \geq \bar{A}$ because, with $F_0'(T(B)) \leq 0$, the right side of (24) is bounded above by $\bar{A}$.

Furthermore, we can generalize (24) to allow for a boundary premium $y \geq 0$. Let $T(B, y)$ denote the horizontal coordinate of the point at which the straight line from (B, y) is tangent to $F_0(\cdot)$. Define $\hat{R}(B, y) \equiv \max_a\{a + F_0'(T(B, y))h(a)\}$ and

$$\mathcal{N} \equiv \{(B, R, y) : 0 \leq R < \hat{R}(B, y), y \geq 0\}.$$

If (B, R, y) is not in $\mathcal{N}$, then, rather than offering a contract to the agent, it is optimal for the firm to collect the value $R + y$ immediately, even in the absence of moral hazard.[36] The closure of $\mathcal{N}$ can be easily shown to be

$$cl(\mathcal{N}) = \{(B, R, y) : T(B, y) \leq W_{gp}^*, 0 \leq R \leq \hat{R}(T(B, y)), y \geq 0\}.$$

A.2 Auxiliary lemma (order of solution curves)

We start with an auxiliary lemma that extends Lemma 2 in Sannikov (2008). For any four numbers a, w, ϕ, and p, define

$$H_a(w, \phi, p) \equiv \frac{a + ph(a) - \phi + \max_c \{p(w - u(c)) - c\}}{\frac{1}{2}r\sigma^2 h'(a)^2}.$$

[35] For example, if $u(c) = \sqrt{c}$, then $\hat{B} = \frac{1}{2}W_{gp}^*$.

[36] It is easy to show that level sets of $\mathcal{N}$, $\mathcal{N}_y \equiv \{(B, R) : 0 \leq R < \hat{R}(B, y)\}$ are nested, with $y' > y$ implying $\mathcal{N}_{y'} \subset \mathcal{N}_y$. Intuitively, with a higher boundary premium y, the set of pairs (B, R) for which the firm would pass on collecting immediately the value $R + y$ is smaller.

Similar to Sannikov (2008), we can express the high-action ODE (10) as

$$F''(W) = -\max_{a\in\mathcal{A}} H_a(W, F(W), F'(W)). \tag{25}$$

Lemma A.1 *Consider two solutions F_1 and F_2 to the high-action ODE that satisfy $F_1(W) \leq F_2(W)$ and $F_1'(W) \leq F_2'(W)$. If at least one of these inequalities is strict, then*

$$F_1'(\tilde{W}) < F_2'(\tilde{W}) \text{ for all } \tilde{W} > W. \tag{26}$$

Proof This proof modifies the proof of Lemma 2 in Sannikov (2008). First, we show (26) in a small neighborhood of W. This holds trivially if $F_1'(W) < F_2'(W)$. If $F_1'(W) = F_2'(W)$, then $F_1(W) < F_2(W)$, in which case

$$F_1''(W) \leq -H_{\tilde{a}}(W, F_1(W), F_1'(W)) < -H_{\tilde{a}}(W, F_2(W), F_2'(W)) = F_2''(W),$$

where $\tilde{a}$ attains $F_2''(W)$ in (25). It follows from $F_1'(W) = F_2'(W)$ and $F_1''(W) < F_2''(W)$ that (26) holds in a small neighborhood of W.

Second, we show (26) for all $\tilde{W} > W$ by contradiction. Suppose (26) does not hold, then there exists a smallest $\hat{W} > W$ at which $F_1'(\hat{W}) = F_2'(\hat{W})$. Since $F_1'(\tilde{W}) < F_2'(\tilde{W})$ for all $\tilde{W} \in (W, \hat{W})$, we have $F_1(\hat{W}) < F_2(\hat{W})$ and again

$$F_1''(\hat{W}) \leq -H_{\tilde{a}}(\hat{W}, F_1(\hat{W}), F_1'(\hat{W})) < -H_{\tilde{a}}(\hat{W}, F_2(\hat{W}), F_2'(\hat{W})) = F_2''(\hat{W}),$$

where $\tilde{a}$ attains $F_2''(\hat{W})$. It follows that $F_1'(\hat{W} - \epsilon) > F_2'(\hat{W} - \epsilon)$ for all sufficiently small $\epsilon > 0$, a contradiction. ■

A.3 Proof of Lemma 2

The following auxiliary lemma starts out by characterizing $x(B, R, y)$ for (B, R, y) on the upper boundary of $cl(\mathcal{N})$.

Lemma A.2 *If $(B, R, y) \in cl(\mathcal{N}) \setminus \mathcal{N}$, then $x(B, R, y) = F_0'(T(B, y)) < 0$.*

Proof If $R = \hat{R}(T(B, y))$, $F(B) = R + y$, and $F'(B) = F_0'(T(B, y))$, then

$$F''(B) = \min_{a,c} \frac{R + y - (a - c + F_0'(T(B, y))(B - u(c) + h(a)))}{\frac{1}{2} r\sigma^2 h'(a)^2} = 0$$

because

$$\begin{aligned} & \min_{a,c}\{R+y-(a-c+F_0'(T(B,y))(B-u(c)+h(a)))\} \\ = \; & \min_{a,c}\{R+F_0(T(B,y))-(a-c+F_0'(T(B,y))(T(B,y)-u(c)+h(a)))\} \\ = \; & R-\hat{R}(T(B,y))=0. \end{aligned}$$

By Lemma 1 in Sannikov (2008), $F''(B)=0$ implies $F''(W)=0$ at all W, i.e., F is a straight line. From definition of $T(B,y)$, F is tangent to F_0+R, i.e., F satisfies condition (19) with $W_{gp}=T(B,y)$. ■

In the remainder of this Appendix, we will denote by $F_{(W,R+y,p)}$ the solution to the high-action ODE, (10), initiated at W with boundary conditions $F(W)=R+y$ and $F'(W)=p$. Here, $R+y$ and p are some generic level and slope of F at W.

Proof of part i of Lemma 2

Given Lemma A.2, it remains to prove part i of Lemma 2 for $(B,R,y)\in\mathcal{N}$. We proceed in three steps. Let

$$K\equiv\max_a\left\{\frac{1+h(a)+W^*_{gp}}{\frac{1}{2}r\sigma^2h'(a)^2}\right\}>0.$$

First, we show that if $F'(B)\geq\bar{A}e^{K(W^*_{gp}-B)}+\bar{A}$, then F is increasing on $[B,W^*_{gp}]$, and hence stays strictly above F_0+R. By contradiction, suppose F is not always increasing but reaches zero slope on $[B,W^*_{gp}]$. Let $\bar{W}$ be the smallest W such that $F'(W)=\bar{A}$. Since $F'(B)>\bar{A}>0$ and F' reaches zero on $[B,W^*_{gp}]$, continuity of F' implies $\bar{W}<W^*_{gp}$. Thus, on $[B,\bar{W}]$ we have $c=0$, $F(W)\geq 0$, and $a\leq\bar{A}\leq F'(W)$. Therefore, at each $W\in[B,\bar{W}]$ we have

$$\begin{aligned} F''(W) &= \min_a\left\{\frac{F}{\frac{1}{2}r\sigma^2h'(a)^2}-\frac{a}{\frac{1}{2}r\sigma^2h'(a)^2}-\frac{h(a)+W}{\frac{1}{2}r\sigma^2h'(a)^2}F'(W)\right\} \\ &\geq \min_a\left\{-\frac{F'(W)}{\frac{1}{2}r\sigma^2h'(a)^2}-\frac{h(a)+W^*_{gp}}{\frac{1}{2}r\sigma^2h'(a)^2}F'(W)\right\} \\ &= -KF'(W), \end{aligned}$$

which implies $\frac{d\log(F'(W))}{dW}=\frac{F''(W)}{F'(W)}\geq -K$. Integrating, we have $\log(F'(\bar{W}))-\log(F'(B))\geq -K(\bar{W}-B)$, or

$$\bar{W}-B\geq\frac{\log(F'(\bar{W}))-\log(F'(B))}{-K}\geq\frac{\log(\bar{A})-\log(\bar{A}e^{K(W^*_{gp}-B)}+\bar{A})}{-K}>W^*_{gp}-B.$$

This contradicts the fact that $\bar{W} < W^*_{gp}$.

Second, we show that $F'(B) = F'_0(T(B,y))$ implies that the solution curve F goes under F_0+R. Indeed, if $F'(B) = F'_0(T(B,y))$, then

$$F''(B) = \min_{a,c} \frac{y+R-(a-c+F'_0(T(B,y))(B-u(c)+h(a)))}{\frac{1}{2}r\sigma^2 h'(a)^2} < 0,$$

because

$$\begin{aligned} & \min_{a,c} y+R-(a-c+F'_0(T(B,y))(B-u(c)+h(a))) \\ = \; & \min_{a,c} F_0(T(B,y))+R-(a-c+F'_0(T(B,y))(T(B,y)-u(c)+h(a))) \\ = \; & R-\hat{R}(T(B,y)) \; < \; 0. \end{aligned}$$

Since F is strictly concave, it remains strictly below the straight line

$$F_0(T(B,y))+R+F'_0(T(B,y))(W-T(B,y))$$

at all $W > B$. Therefore, $F(T(B,y)) < F_0(T(B,y))+R$.

Third, we show the existence of some $x \in (F'_0(T(B,y)), \bar{A}e^{K(W^*_{gp}-B)}+\bar{A})$ such that $F_{(B,R+y,x)} \geq F_0+R$ and $F_{(B,R+y,x)}$ satisfies (19) at some W_{gp}. Let $x \equiv \inf \mathcal{X}$, where

$$\mathcal{X} \equiv \left\{ \tilde{x} \in (F'_0(T(B,y)), \bar{A}e^{K(W^*_{gp}-B)}+\bar{A}) : \; F_{(B,R+y,\tilde{x})}(W) \geq F_0(W) \text{ for all } W \in [B^+, W^*_{gp}] \right\}.$$

The first step of this proof implies that $\mathcal{X}$ is nonempty, so x is well defined. Continuity of the solution curve in $\tilde{x}$ implies $F_{(B,R+y,x)}(W) \geq F_0(W)+R$ for all $W \in [B^+, W^*_{gp}]$, which verifies that $F_{(B,R+y,x)}$ is always weakly above F_0+R. By the second step of this proof, we have $x > F'_0(T(B,y))$. Next, we verify (19) at some W_{gp} and show that $F_{(B,R+y,x)}$ is strictly concave. We consider two cases.

1. $(B,y) \neq (0,0)$. First, take the sequence $\{x-\frac{1}{n}\}_{n=1}^{\infty}$, which converges to x from below. Since $x-\frac{1}{n} \notin \mathcal{X}$, $F_{(B,R+y,x-\frac{1}{n})}$ goes under F_0+R. Let W_n be the smallest point in $[B^+, W^*_{gp}]$ such that $F_{(B,R+y,x-\frac{1}{n})}(W_n) \leq F_0(W_n)$. By Lemma A.1, the curves $F_{(B,R+y,x-\frac{1}{n})}$ are ordered, i.e., $W_{n+1} \geq W_n$. The sequence $(W_n)_{n=1}^{\infty}$, thus, converges to some $W_{gp} \in [B^+, W^*_{gp}]$. Taking the limit in $F_{(B,R+y,x-\frac{1}{n})}(W_n) \leq F_0(W_n)+R$ yields $F_{(B,R+y,x)}(W_{gp}) \leq F_0(W_{gp})+R$. Because $F_{(B,R+y,x)}$ is always above F_0+R, we have $F_{(B,R+y,x)}(W_{gp}) = F_0(W_{gp})+R$, which shows the value-matching condition in (19).

   Second, $F_{(B,R+y,x)}$ is strictly concave. If $F_{(B,R+y,x)}$ is either a convex function or a straight line, then $x > F'_0(T(B,y))$ implies that $F_{(B,R+y,x)}$ is strictly above F_0+R for all

$W \geq B$, violating $F_{(B,R+y,x)}(W_{gp}) = F_0(W_{gp}){+}R$ at some W_{gp}.

Third, we will show $W_{gp} > 0$. If either $B > 0$ or $B = 0 < y$, then $W_{gp} > B \geq 0$. If $B < 0$, we will show $W_{gp} > 0$ by contradiction. If $W_{gp} = 0$, then the strict concavity of $F_{(B,R+y,x)}$ implies $F'_{(B,R+y,x)}(0) < \frac{F_{(B,R+y,x)}(0)-(B+y)}{0-B} = \frac{-y}{0-B} \leq 0 = F'_0(0)$, which means $F_{(B,R+y,x)}$ goes under $F_0{+}R$ immediately after $W_{gp} = 0$, a contradiction.

Finally, it follows from $W_{gp} > 0$ that $F'_{(B,R+y,x)}(W_{gp}) = F'_0(W_{gp})$. So condition (19) is verified.

2. $(B, y) = (0, 0)$. If $x = 0$ then (19) holds with $W_{gp} \equiv 0$, as $F_0(0) + R = R = y + R$ and $F'_0(0) = 0 = x$.

If $x > 0$, then there exists $\epsilon > 0$ such that $F_{(0,R,x/2)}(W) > F_0(W){+}R$ for all $W \in [0, \epsilon]$ because $F'_{(0,R,x/2)}(0) = x/2 > 0 = F'_0(0)$. Since $x - \frac{1}{n} \notin \mathcal{X}$, $F_{(0,R,x-\frac{1}{n})}$ goes under $F_0{+}R$ on $[\epsilon, W^*_{gp}]$. There is a smallest point $W_n \in [\epsilon, W^*_{gp}]$ such that $F_{(0,R,x-\frac{1}{n})}(W_n) \leq F_0(W_n){+}R$. By Lemma A.1, the curves $F_{(0,R,x-\frac{1}{n})}$ are ordered, i.e., $W_{n+1} \geq W_n$. The sequence $(W_n)_{n=1}^{\infty}$, thus, converges to some $W_{gp} \in [\epsilon, W^*_{gp}]$. It follows from $W_{gp} \geq \epsilon > 0$ that $F'_{(0,R,x)}(W_{gp}) = F'_0(W_{gp})$, which verifies (19).

Moreover, $F_{(0,R,x)}$ is strictly concave because

$$F''_{(0,R,x)}(0) = \min_{a,c} \frac{R - (a - c + x(0 - u(c) + h(a)))}{\frac{1}{2} r\sigma^2 h'(a)^2} < 0,$$

which follows from

$$\begin{aligned} & \min_{a,c} R - (a - c + x(0 - u(c) + h(a))) \\ \leq\ & \min_{a} R - (a - 0 + x(0 - 0 + h(a))) \\ =\ & R - (1 + x)\bar{A} \ < \ 0. \end{aligned}$$

Proof of part *ii* of Lemma 2

We show that $x(B, R, y)$ is a continuous function on $cl(\mathcal{N})$. By contradiction, suppose x is discontinuous at some $(B^0, R^0, y^0) \in cl(\mathcal{N})$. Then, there exists $\epsilon > 0$ and a sequence $(B_n, R_n, y_n)_{n=1}^{\infty} \to (B^0, R^0, y^0)$ such that $|x_n - x(B^0, R^0, y^0)| \geq \epsilon$ for all n, where $x_n \equiv x(B_n, R_n, y_n)$. Because $(x_n)_{n=1}^{\infty}$ belongs to the compact set $[F'_0(T(B_n, y_n)), \bar{A}e^{K(W^*_{gp}-B_n)}{+}\bar{A}] \subseteq [F'_0(W^*_{gp}), \bar{A}e^{K(W^*_{gp}-\min_n\{B_n\})}{+}\bar{A}]$, and $(W_{gp,n})_{n=1}^{\infty}$ belongs to the compact set $[0, W^*_{gp}]$, there is a subsequence $(B_{n_k}, R_{n_k}, y_{n_k})_{k=1}^{\infty}$ such that $(x_{n_k})_{k=1}^{\infty}$ converges to some limit x_∞ and $(W_{gp,n_k})_{k=1}^{\infty}$ converges to some limit $W_{gp,\infty}$. Now we show that $F_{(B,R+y,x_\infty)}$ satisfies condition (19) at $W_{gp,\infty}$. By the continuity of F, taking the limit $k \to \infty$ in $F_{(B_{n_k},R_{n_k}+y_{n_k},x_{n_k})}(W) \geq F_0(W){+}R$ shows that $F_{(B,R+y,x_\infty)}(W)$ is always above $F_0(W)+R$. Similarly, by the continuity of F and F', taking the limit $k \to \infty$ in

$$F_{(B_{n_k},R_{n_k}+y_{n_k},x_{n_k})}(W_{gp,n_k}) = F_0(W_{gp,n_k})+R \quad \text{and} \quad F'_{(B_{n_k},R_{n_k}+y_{n_k},x_{n_k})}(W_{gp,n_k}) = F'_0(W_{gp,n_k}),$$

shows that

$$F_{(B,R+y,x_\infty)}(W_{gp,\infty}) = F_0(W_{gp,\infty})+R \quad \text{and} \quad F'_{(B,R+y,x_\infty)}(W_{gp,\infty}) = F'_0(W_{gp,\infty}).$$

This implies x_∞ is equal to $x(B^0, R^0, y^0)$. But $|x_\infty - x(B^0, R^0, y^0)| \geq \epsilon$, a contradiction.

Proof of part *iii* of Lemma 2

We now show that $R_1 < R_2$ implies $x(B, R_1, y) > x(B, R_2, y)$. By definition of $x(B, R_2, y)$, the solution $F_{(B,R_2+y,x(B,R_2,y))}$ pastes smoothly with F_0+R_2 at some $W > B$. Denote this W by $W_{gp,2}$. Lemma A.1 implies

$$F_{(B,R_1+y,x(B,R_2,y))}(W) - (R_1 + y) < F_{(B,R_2+y,x(B,R_2,y))}(W) - (R_2 + y) \text{ for all } W > B.$$

In particular, at $W_{gp,2} > B$ we have

$$F_{(B,R_1+y,x(B,R_2,y))}(W_{gp,2}) - (R_1 + y) < F_{(B,R_2+y,x(B,R_2,y))}(W_{gp,2}) - (R_2 + y) = F_0(W_{gp,2})-y,$$

which means the curve $F_{(B,R_1+y,x(B,R_2,y))}$ goes under F_0+R_1. Since the curve $F_{(B,R_1+y,x(B,R_1,y))}$ must stay above F_0+R_1, it follows from Lemma A.1 that $x(B, R_1, y) > x(B, R_2, y)$.

The proof of $x(B, R, y_1) > x(B, R, y_2)$ for $y_1 < y_2$ is similar, hence it is omitted.

QED

A.4 Proof of Proposition 2

i. **(Region of no contract)** We need to show the sign of $x(B, R, 0)$ for all $(B, R, 0) \in cl(\mathcal{N})$. We proceed in three steps. First, we consider $B < 0$, then $B = 0$, and finally $B > 0$.

   First, we show $x(B, R, 0) > 0$ for all $B < 0$ and $R \in [0, \bar{A})$. Indeed, $B < 0$, $y = 0$, and $R < \bar{A} = \hat{R}(B, 0)$ imply $(B, R, 0) \in \mathcal{N}$. Thus, as shown in the proof of part i of Lemma 2, $x(B, R, 0) > F'_0(T(B, 0))$. But $B < 0$ implies $F'_0(T(B, 0)) = F'_0(0) = 0$. Thus,

$$x(B, R, 0) > 0 \text{ for all } B < 0. \tag{27}$$

Second, we show $x(B,R,0) \geq 0$ for $B=0$. In particular, there exists a $\bar{\boldsymbol{R}} \in (0,\bar{A})$ such that

$$x(0,R,0) \begin{cases} >0, & \text{if } R<\bar{\boldsymbol{R}}; \\ =0, & \text{if } R \geq \bar{\boldsymbol{R}}. \end{cases} \tag{28}$$

If R approaches $\bar{A}$, then $F_{(0,R,0)}$ has near-zero curvature. Indeed,

$$F''_{(0,R,0)}(0) = \min_{a,c} \frac{R-(a-c+0(0-u(c)+h(a)))}{\frac{1}{2}r\sigma^2 h'(a)^2} = \min_{a\in\mathcal{A}} \frac{R-a}{\frac{1}{2}r\sigma^2 h'(a)^2} \geq -\frac{\bar{A}-R}{\frac{1}{2}r\sigma^2 h'(0)^2},$$

which converges to 0 as R approaches $\bar{A}$ from below. With near-zero curvature, $F_{(0,R,0)}$ cannot return to $F_0{+}R$ at any $W_{gp}>0$. We thus define $W_{gp}=0$, which implies $x(0,R,0)=0$. Define $\bar{\boldsymbol{R}} \equiv \inf_R\{R : x(0,R,0)=0\}$. By continuity, $x(0,\bar{\boldsymbol{R}},0)=0$. We have $\bar{\boldsymbol{R}}>0$ because $x(0,0,0)>0$. Clearly, for $R<\bar{\boldsymbol{R}}$, $x(0,R,0)\neq 0$. Since $x(0,R,0)\geq 0$, we have $x(0,R,0)>0$ for $R<\bar{\boldsymbol{R}}$.

Third, we consider $x(B,R,0)$ for $B>0$. There are two cases:

(a) $R \geq \bar{\boldsymbol{R}}$. With $B>0$, Lemma A.1 implies that, with any $p\geq 0$, $F_{(B,R,p)}(W) > F_{(0,R,0)}(W)$ for all $W\geq B$. Since $F_{(0,R,0)} \geq F_0{+}R$, the solution curve $F_{(B,R,p)}$ cannot return to $F_0{+}R$ at any $W_{gp}\geq B$. Therefore, $x(B,R,0)<0$ whenever $B>0$ and $R\geq \bar{\boldsymbol{R}}$. Defining $\bar{B}(R)\equiv 0$ for all $R\geq\bar{\boldsymbol{R}}$, with (27) and (28), we have shown (22) for $R\geq\bar{\boldsymbol{R}}$.

(b) $R<\bar{\boldsymbol{R}}$. By (28), $x(0,R,0)>0$. Lemma A.2 implies $x(\hat{B},R,0)=F_0'(T(\hat{B}))<0$, where $\hat{B}>0$ satisfies $\hat{R}(T(\hat{B}))=R$. Since x is continuous, the intermediate value theorem implies that $x(B,R,0)=0$ for some $B\in(0,\hat{B})$. Let $\bar{B}(R)$ denote the smallest such B. Clearly, $x(B,R,0)>0$ for all $B\in(0,\bar{B}(R))$. If $B>\bar{B}(R)$, then $F_{(B,R,0)}$ stays above $F_{(\bar{B}(R),R,0)}$ and thus also above $F_0{+}R$, hence $x(B,R,0)<0$. With this $\bar{B}(R)$ for $R<\bar{\boldsymbol{R}}$, we have completed the proof of (22). In particular, with $R=0$ and $\bar{B}(0)=\bar{\boldsymbol{B}}$, (22) implies the last statement of Lemma 1.

Since $\bar{B}(\cdot)$ satisfies $x(\bar{B}(R),R,0)=0$ for all R, continuity of $\bar{B}(\cdot)$ follows from the continuity of x. To show that $\bar{B}(\cdot)$ is strictly decreasing on $[0,\bar{\boldsymbol{R}}]$, pick $0\leq R_1<R_2\leq\bar{\boldsymbol{R}}$. Part *iii* of Lemma 2 (monotonicity in R) implies that $x(\bar{B}(R_1),R_2,0)<x(\bar{B}(R_1),R_1,0)=0$, which, by (22), implies $\bar{B}(R_1)>\bar{B}(R_2)$. Finally, as the inverse of a continuous and strictly decreasing function, $\bar{R}(B)$ is continuous and strictly decreasing on $[0,\bar{\boldsymbol{B}}]$. In particular, $\bar{B}(\bar{\boldsymbol{R}})=0$ implies $\bar{R}(0)=\bar{\boldsymbol{R}}$, and $\bar{B}(0)=\bar{\boldsymbol{B}}$ implies $\bar{R}(\bar{\boldsymbol{B}})=0$.

ii. **(Regions of termination and suspension)** Fix $B\in(0,\bar{\boldsymbol{B}})$. If $R=0$, then we have $x(B,R,0)-\frac{R}{B}=x(B,0,0)>0$, where the inequality follows from (22) because $B<\bar{B}(0)=\bar{\boldsymbol{B}}$. If $R=\bar{R}(B)$, then we have $x(B,R,0)-\frac{R}{B}=0-\frac{\bar{R}(B)}{B}<0$. By the intermediate value theorem, the continuity of $x(B,R,y)$ implies the existence of $R^* \in (0, \bar{R}(B))$ such that $x(B,R,0) - \frac{R}{B} = 0$. Part *iii* of Lemma 2 implies that R^* is unique. For $B = \bar{\boldsymbol{B}}$, we have $x(\bar{\boldsymbol{B}},0,0) = 0$, which implies that $x(\bar{\boldsymbol{B}}, R^*(\bar{\boldsymbol{B}}),0) - \frac{R^*(\bar{\boldsymbol{B}})}{\bar{\boldsymbol{B}}} = 0$ holds with $R^*(\bar{\boldsymbol{B}}) = 0$.

Next, we evaluate the first derivative of the function $R^*(B)$ at $B = 0$ and $B = \bar{\boldsymbol{B}}$.

First, we show $R^{*\prime}(0) = x(0,0,0) > 0$. Differentiation of $Bx(B,R^*(B),0) = R^*(B)$ yields

$$B\frac{dx(B,R^*(B),0)}{dB} + x(B,R^*(B),0) = R^{*\prime}(B). \tag{29}$$

At $B = 0$, (29) becomes $R^{*\prime}(0) = x(0,0,0) > 0$.

Second, we show $R^{*\prime}(\bar{\boldsymbol{B}}) < 0$. If $B = \bar{\boldsymbol{B}}$, then $R^*(B) = 0$ and $x(B,R^*(B),0) = x(B,0,0) = 0$, and so equation (29) reduces to

$$\bar{\boldsymbol{B}}\frac{dx(\bar{\boldsymbol{B}},0,0)}{dB} = R^{*\prime}(\bar{\boldsymbol{B}}). \tag{30}$$

Using

$$\frac{dx(B,R^*(B),0)}{dB} = \frac{\partial x(B,R^*(B),0)}{\partial B} + \frac{\partial x(B,R^*(B),0)}{\partial R}R^{*\prime}(B),$$

equation (30) becomes

$$R^{*\prime}(\bar{\boldsymbol{B}}) = \frac{\bar{\boldsymbol{B}}\frac{\partial x(\bar{\boldsymbol{B}},0,0)}{\partial B}}{1 - \bar{\boldsymbol{B}}\frac{\partial x(\bar{\boldsymbol{B}},0,0)}{\partial R}}.$$

Part *iii* of Lemma 2 implies $\frac{\partial x(\bar{\boldsymbol{B}},0,0)}{\partial R} \leq 0$. To finish the proof, it suffices to show $\frac{\partial x(\bar{\boldsymbol{B}},0,0)}{\partial B} < 0$.

Lemma A.1 implies $x(\bar{\boldsymbol{B}}+\epsilon,0,0) < F'_{(\bar{\boldsymbol{B}},0,0)}(\bar{\boldsymbol{B}}+\epsilon)$. Then,

$$\begin{aligned}\frac{\partial x(\bar{\boldsymbol{B}},0,0)}{\partial B} &= \lim_{\epsilon\to 0}\frac{x(\bar{\boldsymbol{B}}+\epsilon,0,0) - x(\bar{\boldsymbol{B}},0,0)}{\epsilon} \\ &\leq \lim_{\epsilon\to 0}\frac{F'_{(\bar{\boldsymbol{B}},0,0)}(\bar{\boldsymbol{B}}+\epsilon) - 0}{\epsilon} \\ &= F''_{(\bar{\boldsymbol{B}},0,0)}(\bar{\boldsymbol{B}}) \\ &< 0,\end{aligned}$$

where the last inequality follows from the strict concavity of $F_{(\bar{\boldsymbol{B}},0,0)}$.

iii. **(Positive boundary premium from suspension)** If $R = R^*(B)$, then $x(B,R,y^*) = \frac{R+y^*}{B}$ holds with $y^* = 0$ by definition of the function R^*. If $R < R^*(B)$, then with $y = 0$ we have $x(B,R,0) > \frac{R}{B}$ again by definition of R^*. With $y = Bx(B,R,0)$, we have $\frac{y+R}{B} = \frac{Bx(B,R,0)+R}{B} \geq x(B,R,0) > x(B,R,y)$, where the strict inequality follows from part *iii* of Lemma 2 (x is strictly decreasing in y). Since x is continuous in y, the intermediate value theorem implies the existence of $y^* \in (0, Bx(B,R,0))$ such that $x(B,R,y^*) = \frac{R+y^*}{B}$. Part *iii* of Lemma 2 also implies the uniqueness of y^*.

QED

A.5 Auxiliary lemma (positive action)

Next, we provide an auxiliary lemma that verifies that the optimal action is strictly positive whenever the optimal contract is derived from a solution to the high-action ODE (10). This lemma will be useful in the proof of the verification Theorem 2.

Lemma A.3 *For (B,R) such that $\frac{R}{B} \le x(B,R,0)$, let F denote $F^* = F_{(B,R+y^*,\frac{R+y^*}{B})}$. For (B,R) such that $\frac{R}{B} > x(B,R,0) > 0$, let F denote $\tilde{F}$. Then,*

i. F satisfies

$$\min_{c\ge 0}\{F(W) + c + F'(W)(u(c) - W)\} \ge 0 \quad \text{at all } W \ge B. \tag{31}$$

ii. The optimal action a^* is nonzero everywhere along the high-action ODE solution F.

Proof

i. We first show that all tangent lines to F are weakly above F_0, i.e., that for all $W \ge B$ we have

$$F(W) + F'(W)(\tilde{W} - W) \ge F_0(\tilde{W}) \quad \text{for all } \tilde{W} \ge 0. \tag{32}$$

Since F is concave, we have $F(W) + F'(W)(\tilde{W} - W) \ge F(\tilde{W})$ for all $W \ge B$ and all $\tilde{W} \ge B$. This implies (32) for all $\tilde{W} \ge B$ because $F \ge F_0 + R \ge F_0$ on $[B,\infty)$. If $B \le 0$, this is all we need to show. If $B > 0$, we still need to show (32) for $\tilde{W} \in [0,B)$. For any $\tilde{W} < B \le W$, the left side of (32) is increasing in W because F is concave. It is thus sufficient to show

$$F(B) + F'(B)(\tilde{W} - B) \ge F_0(\tilde{W}) \quad \text{for all } \tilde{W} \in [0,B).$$

By construction of F, F is flatter at $(B, F(B))$ than the low-action ODE solution through this point, i.e., $F'(B) \le \frac{F(B)}{B}$. (It is strictly flatter if $F = \tilde{F}$.) We thus have:

$$F(B) + F'(B)(\tilde{W} - B) \ge F(B) + \frac{F(B)}{B}(\tilde{W} - B) = \frac{F(B)}{B}\tilde{W} \ge 0 \ge F_0(\tilde{W}).$$

Inequality (31) follows now from (32) by changing the variable $\tilde{W} \in [0, \infty)$ to $u(c) \in [0, \infty)$, where $c = -F_0(\tilde{W})$.

ii. It follows from

$$-\frac{a^* + F'(W)h(a^*) - F(W) + \max_c \{F'(W)(W - u(c)) - c\}}{\frac{1}{2}r\sigma^2 h'(a^*)^2} = F''(W) < 0$$

that $a^* + F'(W)h(a^*) > \min_{c \geq 0} F(W) + c + F'(W)(u(c) - W) \geq 0$, where the weak inequality follows from (31). This implies $a^* \neq 0$.

■

A.6 Proof of Theorem 2

First, we show that, for any (B, R) such that $B \in (0, \bar{\boldsymbol{B}}]$ and $R \leq R^*(B)$, the net profit achieved by an arbitrary incentive compatible contract $(\tau_{tn}, \tau_{gp}, W_{gp}, \{C, A\})$ is at most $F^*(W_0)$, where $W_0 \geq B$ is the agent's initial continuation value in this contract. If $W_0 \geq W^*_{gp}$, then, following Lemma 4 in Sannikov (2008), we can show that the firm's continuation profit is not larger than $R + F_0(W_0) \leq v(W_0)$. If $W_0 < W^*_{gp}$, denote the agent's continuation value under the arbitrary contract by $W_t = W_t(C, A)$, which follows (4) until termination/retirement time $\min\{\tau_{tn}, \tau_{gp}\}$. As in Sannikov (2008), it is without loss of generality to only consider contracts such that $u'(C_t) \geq \gamma_0$ at all t, with which we have that (C_t, A_t) belongs to the compact set $[0, (u')^{-1}(\gamma_0)] \times \mathcal{A}$ at all t. Define

$$G_t \equiv \mathbb{E}\Bigg[r \int_0^{\min\{t, \tau_{tn}, \tau_{gp}\}} e^{-rt}(A_t - C_t)dt + 1_{\{t < \min\{\tau_{tn}, \tau_{gp}\}\}} e^{-rt}(F^*(W_t)) + 1_{\{\tau_{tn} < \min\{t, \tau_{gp}\}\}} e^{-r\tau_{tn}} R + 1_{\{\tau_{gp} < \min\{t, \tau_{tn}\}\}} e^{-r\tau_{gp}} (R + F_0(W_{gp})) \Bigg]. \quad (33)$$

By Ito's lemma, the drift of G_t at all $t < \min\{\tau_{tn}, \tau_{gp}\}$ is

$$re^{-rt}\left(A_t - C_t - F^*(W_t) + F^{*\prime}(W_t)(W_t - u(C_t) + h(A_t)) + r\sigma^2 Y_t^2 \frac{F^{*\prime\prime}(W_t)}{2}\right).$$

Let us show that the drift of G_t is always nonpositive. If $A_t > 0$, then incentive compatibility requires $Y_t = h'(A_t)$. Then the fact that F^* solves the high-action ODE (10) implies that the drift of G is nonpositive. If $A_t = 0$, then (31) and $F^{*\prime\prime} < 0$ imply that the drift of G_t is nonpositive.

It follows that G_t is a bounded supermartingale until the stopping time τ (possibly ∞) defined as the time when the worker either terminates at B (i.e., τ_{tn}), or retires at W_{gp} (i.e., τ_{gp}), or W_t reaches W^*_{gp}. Defining $\tau^*_{gp} \equiv \min\{t : W_t = W^*_{gp}\}$, we have $\tau = \min\{\tau_{tn}, \tau_{gp}, \tau^*_{gp}\}$. If $\tau = \tau_{tn}$, then the firm's continuation profit is $R \leq F^*(B)$. If $\tau = \tau_{gp}$, then the firm's continuation profit is $R + F_0(W_{gp}) \leq F^*(W_{gp})$. If $\tau = \tau^*_{gp}$, then, following Lemma 4 in Sannikov (2008), the firm's continuation profit is at most $R + F_0(W^*_{gp}) \leq F^*(W^*_{gp})$. Therefore, the firm's expected profit at time 0 is less than or equal to

$$\mathbb{E}\Bigg[r\int_0^\tau e^{-rt}(A_t - C_t)dt + 1_{\{\tau=\tau_{tn}\}}e^{-r\tau_{tn}}R + 1_{\{\tau=\tau_{gp}\}}e^{-r\tau_{gp}}\left(R + F_0(W_{gp})\right)$$
$$+1_{\{\tau=\tau^*_{gp}\}}e^{-r\tau^*_{gp}}F^*(W^*_{gp})\Bigg] = \mathbb{E}[G_\tau] \leq G_0 = F^*(W_0).$$

Second, we show that the contract $(\tau_{tn}, \tau_{gp}, W_{gp}(B,R), \{(C,A)\})$ described in the statement of the theorem achieves profit $v(W_0)$ for $W_0 \in [B, W_{gp}(B,R)]$. Existence of a weak solution to (7) follows from Engelbert and Peskir (2014). In particular, a solution exists despite the vanishing of the volatility of W_t at B. Defining G_t as in (33), but now specifically for the stated contract, we have from Ito's lemma that the drift of G_t at all $t < \tau_{gp}$ is

$$re^{-rt}\left(A_t - C_t - F^*(W_t) + F^{*\prime}(W_t)(W_t - u(C_t) + h(A_t)) + r\sigma^2 h'(A_t)^2 \frac{F^{*\prime\prime}(W_t)}{2}\right)$$

if $W_t > B$, and

$$re^{-rt}\left(-C_t - L^*(W_t) + L^{*\prime}(W_t)(W_t - u(C_t))\right)$$

if $W_t = B$. If $W_t > B$, the drift of G_t is zero because F^* solves the high-action ODE (10). If $W_t = B$, the drift of G_t is zero because

$$-C_t - L^*(W_t) + L^{*\prime}(W_t)(W_t - u(C_t)) = -(R + y^*(B,R)) + \frac{R + y^*(B,R)}{B}(B - 0) = 0.$$

It follows that G_t is a bounded martingale until the stopping time τ_{gp}. At τ_{gp}, the agent is retired and the firm's continuation profit is equal to $R + F_0(W_{gp}) = v(W_{gp})$. Therefore, the firm's expected profit at time 0 is equal to

$$\mathbb{E}\left[\int_0^{\tau_{gp}} e^{-rt}(A_t - C_t)dt + e^{-r\tau_{gp}}v(W_{gp})\right] = \mathbb{E}[G_{\tau_{gp}}] = G_0 = v(W_0).$$

QED

References

Atkeson, A. and R. E. Lucas, Jr. (1995). Efficiency and equality in a simple model of efficient unemployment insurance. *Journal of Economic Theory 66*(1), 64–88.

Bou-Rabee, N. and M. C. Holmes-Cerfon (2020). Sticky brownian motion and its numerical solution. *SIAM Review 62*(1), 164–195.

Clementi, G. L. and H. Hopenhayn (2006). A theory of financing constraints and firm dynamics. *Quarterly Journal of Economics 121(1)*, 229–265.

DeMarzo, P. and Y. Sannikov (2006). Optimal security design and dynamic capital structure in a continuous-time agency model. *The Journal of Finance 66*(6), 2681–2724.

Edmans, A. and X. Gabaix (2016). Executive compensation: A modern primer. *Journal of Economic Literature 54*(4), 1232–87.

Engelbert, H. and G. Peskir (2014). Stochastic differential equations for sticky brownian motion. *Stochastics 86*(6), 993–1021.

Fong, Y.-f. and J. Li (2017). Relational contracts, limited liability, and employment dynamics. *Journal of Economic Theory 169*, 270–293.

Fuchs, W. (2007). Contracting with repeated moral hazard and private evaluations. *American Economic Review 97*(4), 1432–1448.

Halac, M. (2012). Relational contracts and the value of relationships. *American Economic Review 102*(2), 750–79.

Harrison, J. M. and A. J. Lemoine (1981). Sticky brownian motion as the limit of storage processes. *Journal of Applied Probability*, 216–226.

He, Z. (2009). Optimal executive compensation when firm size follows geometric brownianmotion. *Review of Financial Studies 22*, 859–892.

Jensen, M. C. and K. J. Murphy (1990). Performance pay and top-management incentives. *Journal of Political Economy 98*(2), 225–264.

Jenter, D. and K. Lewellen (2020). Performance-induced ceo turnover. *Review of Financial Studies, forthcoming.*

Lamadon, T. (2016). Productivity shocks, long-term contracts and earnings dynamics. *Working Paper, University of Chicago*.

Levin, J. (2003). Relational incentive contracts. *American Economic Review 93*(3), 835–857.

Lise, J., C. Meghir, and J.-M. Robin (2016). Matching, sorting and wages. *Review of Economic Dynamics 19*, 63–87.

MacLeod, W. B. and J. M. Malcomson (1989). Implicit contracts, incentive compatibility, and involuntary unemployment. *Econometrica: Journal of the Econometric Society*, 447–480.

Mortensen, D. T. and C. A. Pissarides (1994). Job creation and job destruction in the theory of unemployment. *Review of Economic Studies 61*(3), 397–415.

Phelan, C. (1995). Repeated moral hazard and one-sided commitment. *Journal of Economic Theory 66*(2), 488–506.

Piskorski, T. and M. M. Westerfield (2016). Optimal dynamic contracts with moral hazard and costly monitoring. *Journal of Economic Theory 166*, 242–281.

Pissarides, C. A. (1985). Short-run equilibrium dynamics of unemployment, vacancies, and real wages. *American Economic Review 75*(4), 676–690.

Sannikov, Y. (2008). A continuous-time version of the principal-agent problem. *Review of Economic Studies 75*(3), 957–984.

Spear, S. E. and C. Wang (2005). When to fire a CEO: Optimal termination in dynamic contracts. *Journal of Economic Theory 120*(2), 239–256.

Stiglitz, J. E. and A. Weiss (1983). Incentive effects of terminations: Applications to the credit and labor markets. *American Economic Review 73*(5), 912–927.

Szydlowski, M. (2019). Incentives, project choice, and dynamic multitasking. *Theoretical Economics 14*(3), 813–847.

Wang, C. (1995). Dynamic insurance with private information and balanced budgets. *Review of Economic Studies 62*(4), 577–595.

Wang, C. (2011). Termination of dynamic contracts in an equilibrium labor market model. *Journal of Economic Theory 146*(1), 74–110.

Zhu, J. Y. (2013). Optimal contracts with shirking. *Review of Economic Studies 80*(2), 812–839.

Zhu, J. Y. (2018). Better monitoring... worse productivity? *Working Paper*.

Zilberman, E., V. Carrasco, and P. Hemsley (2019). Risk sharing contracts with private information and one-sided commitment. *Economic Theory 68*(1), 53–81.